Annex 1. Summary of types of automated technology discussed in the literature

In Chapters 3 and 4, we provide a high-level overview of the impacts of using automated technology, as well as the factors that support and hinder the use of these tools. In the table below, we provide a more detailed overview of the specific technologies discussed in the literature, for example, software or algorithms.

In multiple instances in the literature, specific technologies were named or mentioned but no detail on what they do or any information on what does and does not work in their use were provided. These have been included in the table as a record that the technology is available so it can be used as a resource for ECDC on the types of approaches available, but we are unable to reflect on how they can be used. In addition, the information provided in this table includes only that which was presented in the reviewed articles. It may be that the technologies listed here have other strengths, limitations and gaps/needs in their use that were not mentioned.

A further online resource is available, SR toolbox, which provides a comprehensive list of the types of tools that are available for evidence synthesis (both automated and non-automated). Users can filter tools by the type of review and stage of review. The resource can be accessed here: http://systematicreviewtools.com/index.php.

Table 1. Overarching approaches to the automation of evidence synthesis

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Text mining (general)	A form of artificial intelligence that uses linguistics, statistics, computer science, natural language processing, machine learning and data mining to convert unstructured text to a structured form that can be extracted and analysed (17, 18, 23, 61-63). It can be used for automatic searches, prioritise articles for screening based on relevancy, screen articles, assess quality, extract quantitative information triangulate data (17, 18, 23, 65, 68, 79).	Literature search, screening, quality assessment, extraction, analysis and write-up (17, 21, 23, 38, 44, 61-68, 79)	Ability to draft text in multiple languages (38). Identify new conclusions that are not identified in non-systematic reviews (62). Reduces the burden of reviews while maintaining high recall and demonstrate good performance (63-66, 68). May be particularly useful for simple reviews and can be used as an additional tool for complex reviews (63). Analysis of topics allows systematic exploration of what types of topics are published in the literature (64). Allows for the identification of current and future research trends (79). Prevents duplications of reviews being published by identifying trends and subjects already covered in existing research (79).	Algorithms are not subjective (17). Unsupervised/active learning models do not need any training (64). Can be adaptable and flexible to add further features (64). Able to rank articles by risk of bias accurately (65).	Interpretation cannot be completely objective (although this is the case for all statistical analysis) (17). Interpretation can also be influenced by whether the algorithm understands meanings of words and associations between (and weight of) variables in the correct way (17, 62). Large body of text is needed to ensure analyse are robust (17). Format of text is a barrier for using text mining (e.g. images cannot be analysed) (17). Lack of transparency with algorithms (can be seen as “black box”) (17). Most text mining tools are limited to use in English language articles (17). Expertise needed in understanding machine learning and statistics (29). Some models only use simple approaches to combining relevant metrics (64). Some models are limited to searching one or few literature databases (64). Performance can be reduced if tool is used on title/abstract rather than full text (65). Literature searching may be less sensitive than traditional approaches (63). May not improve search evaluation time or identification of irrelevant articles (63). Manual input is still needed (63). Availability of a large number of tools makes it difficult to select the most appropriate one and effort may be put into training researchers in a non-optimal tool (63). Ideal size of the training set is not known (63).	Still a fairly new technology and reliability has not been firmly demonstrated yet (17, 23, 64). May not yet meet robust requirements for acceptance by journals (38). Resource costs needed to train researchers in using the technology (63).

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Machine learning (general)	A type of artificial intelligence that uses computational methods (18, 47) to learn over time, improving the accuracy of a task over time without the need for explicit programming to do so. There are two ways in which machine learning can operate, by using supervised or unsupervised learning. Supervised machine learning requires human input to initially train an algorithm on how to perform a task (48, 49). Unsupervised learning does not need to be trained by a human. Rather, it learns using patterns in text and data and mimics these patterns (26). For both types of learning, the algorithm continues to learn and improve over time.	Literature searches, screening, quality assessment, extraction, analysis (18, 29, 37, 39, 44, 47-56).	Reduces the workload of systematic reviews, e.g. reduce number of articles for manual screening or replace a second reviewer (47-50, 52, 54, 56). One study estimated that 75% fewer articles needed manual screening (50) and another that 88-98% of labour could be reduced during the second screening phase (54).	Although further research is needed on efficacy and accuracy, early studies suggest machine learning algorithms are accurate compared to traditional methods (18, 47, 50, 53, 59). Machine learning algorithms can learn over time to become more accurate and replicate human decisions (21, 39, 48). Semantic features can be created automatically, creating further time-savings for the researcher (48). Flexible, adaptable and can deal with complex wording (48). Transparency is improving over time, e.g. some tools can provide reasons for excluding articles (47).	Human input still needed (37, 49) For example, pre-processing steps can be needed to ensure articles are in a format that can be reviewed by the algorithm (e.g. lower case, removing stop words) (18) Researcher input is also needed to confirm machine learning decisions and to train the algorithm (37). Certain algorithms or classifiers may demonstrate less than ideal accuracy (e.g. low recall rate) (18, 39, 48-50, 54) Some may need optimisation to boost performance (53). Compresses time needed for the entire review but does not significantly reduce the amount of time needed for individual researchers (37). Cannot be used to take over completely more cognitive/intellectual tasks and independent decision-making (37). Some models cannot screen based on title alone, so articles without abstracts require manual review (50). Performance is better when reviewing full-texts rather than title/abstracts only (47). Training set needs to be well balanced between papers which do or do not meet the inclusion criteria (47). Algorithms tends to be seen as black boxes and it may be difficult to record why decisions were made (47). The best machine learning algorithm to use remains unclear and the algorithm selected can influence the results of a review (47). Unclear if complex articles can be reviewed effectively (47).	Further research into efficacy and accuracy of machine learning tools, including consistency in studies assessing performance to compare across studies (18, 29). Limited availability of annotated training datasets which take time, expertise and money to develop (18). Improved training sets are needed that are robust and objective, based on a large sample of high quality articles and be up to date (18). Lack of guidance for reporting the use of machine learning for evidence synthesis, particularly for non-clinical reviews, leading to a lack of consistency in reporting and difficulty comparing across studies (18). A combination of machine learning and manual input may be optimal (47). Improved (standardised) article indexing is needed (53).
Natural Language Processing	A computational approach using artificial intelligence (combined with linguistics and computer science) to analyse and interpret text). It can perform basic tasks, such as counting word frequencies, to more complex activities, such as classifying and understanding text (21, 29, 58, 59).	Literature search, screening, quality assessment, extraction, analysis (21, 29, 44, 51, 55, 58, 59).	—	—	Some human input still needed to finetune the algorithm (29). Some algorithms do not demonstrate good performance, e.g. when data is scarce or unlabelled (59, 143). Require large datasets for training which limits when the technology can be used (143).	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Neural Networks	A group of computing algorithms that can identify similarities and relationships between words across articles (70) by mimicking the way neural networks within the human brain work. They can take the form of, for example, deep-learning (have more than three networks) or recurrent neural (enable temporal analysis) networks (22, 71).	Screening (22, 70, 72).	Reduces workload for researchers while increasing yield (70). One study of deep learning neural networks noted manual screening burden was reduced by 50% (72).	Demonstrated good performance and sensitivity in identifying relevant articles (59, 72).	Performance is dependent on accuracy of text indexation and pre-processing of data (70, 72). One study on deep learning neural networks did not outcompete conventional machine learning methods (72).	—

Table 2. Specific approaches to the automation of evidence synthesis

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Automatic web-based question-answering system	Identifies systematic reviews that answer a research question without needing to develop a search query (27).	Develop research questions (27).	—	—	• Initial research needed to identify the optimal system to use (27).	—
Machine learning based document classifier	Alerts researchers to articles that can be included in systematic review updates and identifies those that are likely to be most relevant using annotations (81).	Develop research questions, literature search, screening (81)	• Useful tool for planning, scheduling and allocating resources for updating systematic reviews (81). • Able to recognise that some articles are more important than others (81).	• Demonstrates good performance without requiring significant resources to monitor newly identified articles (81). • The recall rate[1] can be set depending on the resources available to conduct the review (81).	• Performance varied depending on research topic (81).	Further work needed to improve performance (81). Senior/experienced researcher with knowledge of the topic is needed to decide when to conduct the review update (81).
Papyrus	Uses natural language processing to automatically search Medline (24).	Develop research questions, literature search (24).	• (Quickly) identifies additional relevant articles in comparison to traditional methods (24). • Can help inform the focus of future reviews by identifying gaps in evidence (24).	—	• Some manual input required to review some abstracts (24). • Some relevant articles may be missed (24). • May be over-inclusive, e.g. include articles on very rare conditions (24).	Further testing and validation of tool is needed (24).
SWIFT-review	Interactive platform that supports the development of research questions and prioritisation of articles to review by identifying topics over-represented in the literature, and searching and categorising patterns in literature search results. Tool is based on statistical modelling and machine learning (129).	Develop research questions, screening, extraction (32, 43, 44, 129, 130).	—	Free to use (129).	• Could not extract some items (43). • Does not notify researchers when screening can be stopped once the desired recall has been reached (129). • Only works on articles in PubMed (129).	—
Quick Clinical	Search engine for searching clinical research.	Develop research questions, literature search (27, 44).	—	—	• Limited use for systematic reviews as it only searches a small number of databases (27). • Snowballing aspect can retrieve too many articles that are manageable to review manually (27).	Stopping criteria and automatic appraisal system needed to prevent excessive articles being identified during snowballing (27).

[1] The recall rate is the number of articles or items correctly identified by a model (true positives) divided by the total number of relevant articles (true positives + false negatives). For example: if a model identifies 8 articles out of a possible 10, the recall rate is 0.8. This provides a measure of how accurate a model is in identifying the correct articles.

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Microsoft Academic Search	Searches the reference lists of articles to snowball for further relevant articles for a systematic review update (132).	Literature search (132)	—	Demonstrates good recall over 85% (132).	• It is not accurate enough to completely automate the literature search for a systematic review update (132).	—
Arden syntax	Uses automated query construction and arden syntax (a language used to represent medical and clinical information) to search for evidence online (150).	Literature search (150)	—	Demonstrated good performance (150).	• Manual effort is needed to investigate operators and control structures (150).	Technical knowledge is required on how to use the tool (150).
Fast Correlation-Based Filter (FCBF)	Algorithm to identify additional relevant studies (154).	Literature search (154)	—	Effective at identifying additional clinical evidence (154).	—	—
TheoryOn	—	Literature search (21)	—	—	—	—
LitBaskets	—	Literature search (21)	—	—	—	—
LitSonar	—	Literature search (21)	—	—	—	—
Automated daily searches (unnamed)	—	Literature search (83, 84)	—	—	—	—
Automated literature search (unnamed)	Automatically produces a search query (155).	Literature search (155)	• Reduced time needed to search for literature while minimising missed articles (155).	—	—	—
S3EF	Uses supervised and unsupervised learning to review unlabelled text and data (25).	Literature search (25)	—	—	—	—
TF-IDF	Ranks documents when retrieving literature (25).	Literature search (25)	—	—	—	—
DSSM	Calculates the relevancy between a question and a document (25).	Literature search (25)	—	—	—	—
CDDSM	Calculates the relevancy between a question and a document (25).	Literature search (25)	—	—	—	—
DRMM	Calculates the relevancy between a question and a document (25).	Literature search (25)	—	—	—	—
MatchPyramid	Calculates the relevancy between a question and a document (25).	Literature search (25)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
KNRM/Conv-KNRM	Calculates the relevancy between a question and a document. Conv-KNRM can capture more subtle differences in text (25).	Literature search (25)	—	—	—	—
Automatic search algorithms	Algorithm that can improve the search query (automatically or guide researchers to improve it), cluster documents that are similar and by how well they match criteria, collate searches from multiple databases and allow researchers to use one search query across multiple databases (27).	Literature search (27)	—	—	—	—
QUOSA	Allows one search query to be used across multiple databases and collates the results together (27).	Literature search (27)	—	—	—	—
Turning Research into Practice	Uses multiple search query fields and knowledge about evidence-based medicine to conduct a more precise search (27).	Literature search (27)	—	—	• Effectiveness for systematic reviews is not yet demonstrated (27).	—
Sherlock	Search engine for trial registries (27).	Literature search (27, 44)	—	—	• Limited to only searching clinicaltrials.gov (27).	—
Metta	Search engine for use for systematic reviews (27).	Literature search (27, 44)	—	—	• Not publicly available (27).	—
Polyglot Search	—	Literature search (32)	—	—	—	—
Trial2rev	—	Literature search (32)	—	—	—	—
SRA	—	Literature search (32)	—	—	—	—
RCT	Researcher conducts a review as usual and the RCT machine learning algorithm removes study designs not matching the inclusion criteria (i.e not RCTs) (22, 33).	Literature search (22, 33, 44)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
BibExcel	Tool for measuring frequency of selected terms, generating a list of citations containing certain terms (35).	Literature search (35)	• Particularly useful for identifying articles that were not relevant (35).	• Quick and effective analysis (35).	• Time required to analyse the results of the automatic searches (35). • Pre-processing of citations needed to be analyse some text fields (35). • Time and resources needed to effectively use the tool (35).	Prior experience of using the tool was a benefit (35).
AntConc	Identifies sequences of certain numbers in large volumes of text (35).	Literature search (35)	—	• Quick and effective analysis (35).	• Time required to analyse the results of the automatic searches (35). • Pre-processing of citations needed to be analyse some text fields (35). • Time and resources needed to effectively use the tool (35). Access limited by some institutional firewalls (35)	Prior experience of using the tool was a benefit (35).
Voyant Tools	Collection of tools that can visualise the proximity and frequency of words (35).	Literature search (35)	—	• Quick and effective analysis (35).	• Time required to analyse the results of the automatic searches (35). • Pre-processing of citations needed to be analyse some text fields (35). • Time and resources needed to effectively use the tool (35).	Prior experience of using the tool was a benefit (35).
Termine	Automatic search term recognition by making linguistic associations from text (35).	Literature search (35, 44)	—	• Quick and effective analysis (35).	• Time required to analyse the results of the automatic searches (35). • Pre-processing of citations needed to be analyse some text fields (35). • Time and resources needed to effectively use the tool (35).	Prior experience of using the tool was a benefit (35).
Push search model	Model that can automate literature searches in different literature databases and notifies researchers when there are new articles to screen (39).	Literature search (39)	• Allow searches in databases where automation is not usually possible (e.g. for unpublished articles) (39).	—	• Not compatible with all literature databases (39). • Human input still required to search for unpublished literature (39).	—
Epistemonikos	Allows one search string to be used across multiple literature databases (41).	Literature search (41)	—	—	—	—
Health Database Advanced Search	Allows one search string to be used across multiple literature databases (41).	Literature search (41)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
SCOPUS method	Semi-automated approach to snowballing. The tool is shown relevant 20 articles and it downloads the reference lists, removed duplicates and merged the list with the project database. New references are then reviewed manually for relevance (112).	Literature search (112)	• Avoids reviewing duplicate references from the primary search (112). • Saves 63% of researcher time (taking 3 hours compared to 8 for gold standard method) (112).	• Performed with equal validity as gold standard method (112).	• Requires a paid subscription (112). • Does not include refence lists of Cochrane reviews (112).	—
ParsCit	Algorithm used for snowballing that can identify references and convert it into text that can be searchable online (114).	Literature search (114)	—	• Open source (114) • Demonstrated good performance in retrieving citations (114).	—	—
Automated Full Search	Automated literature search (149).	Literature search (149)	—	—	—	—
Automated search strategy (unnamed)	Automatically edits literature searches up to 5 times to make the search increasingly restrictive (as long as there are at least 50 search hits) (141).	Literature search (141)	—	• Demonstrated good efficacy (141).	• Performance is impacted if abstracts are not available (141) • Relied on MeSH terms only (141). • Performance may be impacted by researcher skills in developing initial literature search (141).	-
Visual data mining	Connects similar items between articles to allow researchers to identify other relevant documents (23).	Literature search (23)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Lingo3G	Automatic clustering tool which analyses text distribution in short documents to identify documents using similar clusters of words (35, 89).	Literature search, analysis (35, 44, 88, 89)	• Quick, saving an average of 33% of time and clusters can be generated in less than one second (35, 88, 89). • Can capture topics not identified by researchers (88, 89). • Outlines dominant themes and provides a focus for further exploration of a large dataset (89). • Aid development of search strategies, e.g. identify new terms to use, remove irrelevant terms (89). • Can be used to check the quality of manual coding (89).	• Effective analysis, demonstrating good precision and clustering (35, 88). It can perform as well as a human researcher (88). • Can be used to detect 19 languages (88). • Range of flexible settings (88). • Articles can be manually re-assigned to different clusters (89).	• Time required to check (and edit if needed) the results of the automatic searches and clustering – cannot fully replace humans (35, 88, 89). • Pre-processing of citations needed to be analyse some text fields (35). • Time and resources needed to effectively use the tool (35). • May have difficulties in distinguishing different study types and specific populations (88, 89). • There can be multiple ways of interpreting the clusters due to differences in language meanings (89). • Recall varies across different topics (89). • May not provide detailed coding of articles (89). • May not identify all categories or research gaps (89). • Performance relies on the clarity of text in the title/abstract (89).	—
Crawler classification linked to Wikipedia	Crawler classification links with Wikipedia to find medical articles. Semantic classification is then used to categorise articles. The tool can also assess the quality of articles (140).	Literature search, quality assessment (140)	—	• Using the connected nature of Wikipedia pages allows for better classification than other machine learning methods (140). • Can assess the quality of articles to not include information from low quality articles (140).	• Accuracy is lower for shorter articles (140). • Accuracy is influenced by the classification with Wikipedia which is not always correct (e.g. mixing human and animal topics) (140)	—
SyRF	Automatically retrieve articles from search engines and can integrate other automation tools (28).	Literature search, screening (28, 32)	—	—	—	—
Python based algorithms	Extract meta-analysis data from articles using a customised search query. Natural Language Processing and unsupervised machine learning are used to identify frequency and distribution of words	Literature search, screening (29, 92, 103)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
	and identify the most relevant articles for screening (29, 103).					
Thalia	Machine learning based tool that indexes new PubMed articles on a daily basis based on topic (e.g. drug, disease) (33).	Literature search, screening (33, 44)	—	—	—	—
Bibliography BOT (BiBot)	Uses keywords and natural language processing to identify articles from PubMed and interpret key words in abstracts (110)	Literature search, screening, analysis (32, 110)	• Saves time – search took 3 hours compared to 4 hours a month for 4 months manually (110). • Reduced the need for manual input 4.4 fold compared to manual search (110).	• Demonstrates good reliability in retrieving relevant articles (110). •	• Some relevant articles may have been missed (110). • Articles could not be rated as the tool cannot evaluate PRISMA checklist items (110).	• Combining BiBot with manual approaches may achieve best results (110).
AutoLit	Platform that uses nested knowledge to quickly identify, bring together and analyse data (96).	Literature search, screening, analysis (96)	• Quick and streamlined (searching and excluding articles took less than 1 minute, screening of other articles took 2 hours and extraction took under half an hour) (96). • There is a full audit record of the search, screening criteria, screening decisions and organisation of data, preventing duplication of work (96). • Article and outcome variables are together under the same function which reduces error and supported screening decisions (96).	• Able to identify correct articles (96). • Training not needed to use the tool (96).	—	—
Cochrane Crowd	Crowdsourcing platform allowing individuals to contribute to a review via 'microtasks' (41).	Literature search, screening, extraction, analysis (41).	—	• High sensitivity (99%) (41).	• Algorithm requires each article to be classified multiple times to ensure accuracy (41).	—
CADIMA	-	Literature search, screening, quality assessment, extraction (32, 126)	—	• User friendly, flexible and can be used offline (126). • Demonstrates good performance, including with large datasets, compared to similar tools (126). • Can be used with multiple users (126). • Secure platform for organising articles and encourages researchers to document decision-making (126). • Free to use (126).	• The same researchers need to take part in each screening stage – new researchers cannot be added (126) • Can identify but not automatically remove duplicates (126) • Can only recognise RIS files (126).	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Abstrackr	Machine learning and text mining tool that prioritises articles for screening based on relevancy. It can learn from human screening decisions to improve over time (22, 27, 44, 98, 100, 102, 120).	Screening (22, 23, 27, 32, 43, 44, 98-100, 102, 120)	• Reduced workload for researchers and saves time, particularly in cases where there is low risk of excluding the wrong articles (43, 98-100, 102, 120). More time was saved during systematic reviews than rapid/descriptive reviews (100, 120).	• Demonstrates good specificity and low numbers of false negatives (102). • User friendly and easy to use (99). • Trusted (99). • Freely available (120).	• Did not identify all relevant articles (recall may be reduced by 5%) (27, 43, 98-100). One study estimated that 14% of articles were wrongly classified (98). Others estimate 6% of records were missed (120) and false positive rates were 12.6% (100). Greater reduction in manual workload associated with greater number of missed articles (99, 100). • Can overestimate the relevancy of some articles (102). • Requires some expertise in understanding how the tool works to be used effectively (23). • Performance is better with mixed methods and qualitative studies over observational studies and reviews (98). • Difficult to find clear instructions on using the tool and time was spent on troubleshooting issues (102). • A lack of abstract means tool has to use key words selected by researchers which is less reliable (102).	—
Adaboost	A machine learning algorithm using decision trees generated in sequence, based on learning from previous mistakes for use in text classification (30).	Screening (30)	—	• Can be used in combination with other machine learning algorithms and neural networks (30).	• Human input and expertise needed to label data to train algorithm (30). • Training set needs to be developed based on specific topic of focus (30). • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30).	—
ADIT approach		Screening (21)	—	—	—	—
Algorithmic automation to refine Boolean queries	Automation of Boolean search queries (156).	Screening (156)	—	• Information from included and excluded articles can be used to improve the tools accuracy (156).	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
ASR Review	Uses active learning to regularly updates the relevancy of unscreened abstracts by learning from researcher screening (21, 31, 125).	Screening (21, 31, 125)	• Minimises the number of articles that need screening and prioritises those that a most relevant, saving researcher time (21, 31, 125).	• High accuracy for identifying relevant articles (31). • Easy to use graphical interface which allows modification of algorithms (21, 125) • Can be used across different databases (21) • Open source and free to use (125) • Can be used with multiple different machine learning models and new modules can be added by third parties to improve functionality (125) • Can be used with any text source (125)	• Cannot easily provide an error rate (125)	Lack of research into the performance of the tool and benchmarking against other tools (125)
Bibliometric techniques	Automatic article selection and provides overview of trends in research topics (78)	Screening (78)	• Improves screening efficiency (78) • Reduces bias in screening (78) • Organises research trends over time and this can be re-run to produce increasingly specific topics (78) Also gives detailed insights into research topics and trends (78) • Allows for exploration of future research trends (78)	• Limited prior knowledge about the research topic needed to use the tool (78) •	• Often limited to articles in English (78) • There may be bias against certain types of articles, e.g. using a novel method and newer papers (78) • Can include articles that are not relevant (78) • Separation of research topics is not perfect (78)	—
Biomedical Research Article Distiller (BioReader)	Uses text mining to classify articles based on a training set (93)	Screening (93)	—	• Demonstrates good performance when compared to similar tools (93) • Multiple machine learning tools can be used to allow classification of different sizes and complexity of text (93)	• Performance is influenced by the training set, e.g. size and how well it distinguishes relevant and irrelevant articles (93)	—
Boostexter	Group of algorithms that can be used for text categorisation to combine somewhat inaccurate rules into one, accurate rule (151)	Screening (151)	—	• Demonstrates comparable performance to traditional tools (151)	—	Technical knowledge is required to be able to use the tool (151)
Certainty Criterion	Form of active machine learning to support screening (109)	Screening (109)	• Reduces the burden for researchers without impacting on performance (109)	• Is able to identify relevant articles (109)	—	—
Classifiers	Decides what articles to include and exclude (23)	Screening (23)	—	—	—	—
Covidence		Screening (32, 44)	—	—	—	—
DRAGON		Screening (32)	—	—	—	—
Efficient citation assignment	Selects articles for researcher screen (23)	Screening (23)	—	—	—	—
Evidence in Documents, Discovery, and Analysis (EDDA)	Hybrid system to reduce the number of articles that require re-screening (57)	Screening (57)	• Reduces the burden of screening, particularly the second round of re-screening articles – estimates suggest 97% reduction in workload (57)	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
EvoSVM	Supervised machine learning tool for screening (52)	Screening (52)	—	• Demonstrated very high recall in training conditions (52)	—	—
Factorized version of the complement naïve Bayes (FCNB) classifier	Classifies abstracts after being trained as relevant or not (108)	Screening (108)	Manual workload reduced by an average of 36% (108)	• Reaches a 95% recall rate (108) • Interoperable with other tools (108)	• When used alone, it may be less accurate (108)	—
GAPScreener	Text mining tool	Screening (44)	—	—	• Requires some expertise in understanding how the tool works to be used effectively (23)	—
Gradient Boost	A machine learning algorithm using decision trees generated in sequence for use in text classification (30)	Screening (30)	—	• Can be used in combination with other machine learning algorithms and neural networks (30)	• Human input and expertise needed to label data to train algorithm (30) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30)	—
Integrated Network and Dynamical Reasoning Assembler (INDRA)	Machine learning and network visualisation tool that use natural language processing o aggregate results about biological and chemical mechanisms (87)	Screening (87)	• Supports the interpretation of data (87)	—	• There is a risk of excluding relevant articles (87)	—
JBL Sumari	-	Screening (32)	—	—	—	—
K-means	Clustering algorithm (43)	Screening (43)	—	—	—	—
LitSuggest	-	Screening (32)	—	—	—	—
LivSVM	-	Screening (32)	—	—	—	—
Long short-term memory	A machine learning algorithm which reviews text and memorises it as it goes for use in text classification (30)	Screening (30)	—	• Can be used in combination with other machine learning algorithms and neural networks (30)	• Human input and expertise needed to label data to train algorithm (30) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30)	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Mechanical Turk	Crowdsourcing platform using non-researchers to screen articles (157)	Screening (157)	• Rapid screening decisions (157)	• Relatively high screening accuracy compared to manual screening, despite users lack of familiarity with research (157) • Easy to access by a large group of people and easy to compensate users for their time (157) • Range of functions that can be used, e.g. quality control, qualification tests (157) • Quality control tests encourage better quality responses (157) • Inexpensive compared to manual screening costs (157)	• Complex topics require more support from researchers for reviewers (which can require effort from researchers) (157) Minimising the complexity of inclusion criteria may result in a higher yield of articles (157) • Can receive significant amount of malicious (careless or wrong) responses (157) • There can be significant disagreement across reviewers (157) • Did not identify all relevant articles (157)	—
Metagear	-	Screening (32)	—	—	—	—
PARSIFAL	-	Screening (32)	—	—	—	—
Pimiento	Text mining tool	Screening (44)	—	—	• Requires some expertise in understanding how the tool works to be used effectively (23)	—
Pool-based active learning methods	Text mining approach to annotate articles (109)	Screening (109)	• Can be used for reviews on complex topics, including social sciences (109)	• Demonstrated good performance (109)	• Require significant computational cost and memory (109)	—
QDA Miner	-	Screening (43, 102)	• Reduced workload for researchers (43)		• Some issues in uploading some articles due to formatting (102)	—
Random Forest	A machine learning algorithm that creates decision trees to make a prediction (30, 143)	Screening (30, 59)	—	• Shown to be effective (30, 59) • Can be used in combination with other machine learning algorithms and neural networks (30)	• Human input and expertise needed to label data to train algorithm (30) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30)	—
Ranking Prioritisation Systems	Excludes articles that do not meet criteria (23)	Screening (23)	—	—	—	—
RapidMiner	Text mining tool	Screening (44)	—	—	• Requires some expertise in understanding how the tool works to be used effectively (23)	—
ReLiS	—	Screening (32)	—	—	—	—
Research Screener	—	Screening (32)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Reviewer Terms	Researchers develop list of terms that are relevant and one that are not. The tool reviews title and abstracts to calculate the number of relevant and irrelevant terms (20)	Screening (20)	• Analyse large amounts of complex data across different subject areas (20) • Reduced manual screening time while identifying 10x more relevant studies than manual approaches (20)	• Technically easy to use (20)	• Greater reliance on researcher input than some other screening tools • Large amount of computer power needed if large datasets are used (20) • Human input still needed to train tool and respond to information being produced (20) • Large number of articles meant process was still time consuming (20)	Initial training needed to understand how to use the tool effectively (20) Challenges in identifying appropriate metrics to analyse the performance of the tool (20)
Revis	Text mining tool	Screening (44)	—	—	• Requires some expertise in understanding how the tool works to be used effectively (23)	—
RobotAnalyst	Machine learning and text mining approach to title and abstract screening (44)	Screening (44)	—	• User friendly and easy to use (99) • Trusted (99)	• Greater reduction in manual workload associated with greater number of missed articles (99)	—
RobotSearch	Uses neural networks and machine learning to screen articles for key words (33)	Screening (33, 44)	—	—	—	—
RysannMD	Software to annotate biomedical literature (66)	Screening (66)	• Led to time savings (66)	• Demonstrates good sensitivity and specificity (66)	• Manual set up can be time consuming (66)	Important to set up the technology with researchers to ensure they are confident in using it. Led to time savings (66)
Single screening with text mining (unnamed)	Text mining algorithm trained in screening criteria to then automatically screen articles (91)	Screening (91)	• Workload reduced by 60% and reduced costs (91)	—	• Recall rate may not be high enough for use alone (91)	—
Singular Value Decomposition (SVD)	-	Screening (66)	• Led to time savings (66)	• Demonstrates good sensitivity and specificity (66)	• Manual set up can be time consuming (66)	Important to set up the technology with researchers to ensure they are confident in using it. Led to time savings (66)
SMOTE	Algorithm that balances the distribution of relevant and irrelevant articles (66)	Screening (66)	• Led to time savings (66)	• Demonstrates good sensitivity and specificity (66) •	• Manual set up can be time consuming (66)	Important to set up the technology with researchers to ensure they are confident in using it. Led to time savings (66)
SRAHelper	—	Screening (44)	—	—	—	—
SRDP.PRO	—	Screening (32)	—	—	—	—
StART	—	Screening (32)	—	—	—	—
SWIFT-Active	—	Screening (43, 44)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
SYRIAC	Support vector machine based tool that automatically collects systematic review inclusion criteria and prioritises articles for review (144)	Screening (144)	—	• Demonstrated good performance in prioritising articles for review (144)	• Some manual correction of references is needed (144)	—
Waterloo CAL	Uses active learning to organise articles by relevancy (85)	Screening (85)	• Allows for quicker and more reliable screening (85)	—	—	—
Wordstat	—	Screening (43, 102)	• Reduced workload for researchers (43)	—	• Some issues with the format of some articles (e.g. PDFs) (102)	—
WSS95	Uses information provided by researchers on examples of relevant and irrelevant articles to identify the next most relevant article (158)	Screening (158)	—	• Demonstrates similar effectiveness to other machine learning models (158) • Could be applied to a broader range of topics than other machine learning tools (158) • Easy to use (158) • Does not require training of a model (158)	—	—
Bag of Words modelling	—	Screening and data extraction (44)	—	—	—	—
Colandr	—	Screening and data extraction (44)	—	—	—	—
SearchFinding	—	Screening and data extraction (44)	—	—	—	—
EPPI-Reviewer	Cloud based platform which can conduct screening (using active learning), qualitative and quantitative analysis, can categorise data and allows researchers to keep a track of decision-making (45, 90)	Screening, analysis (43-45, 90)	—	• Same file can be used by multiple researchers (45)	• Requires some expertise in understanding how the tool works to be used effectively (23) • Requires paid subscription (45) • One study estimated only 40% of articles were correctly identified for inclusion (90)	—
Latent Dirichlet allocation	Form of active text mining that identifies common topics across abstracts (66, 69, 92, 109)	Screening, analysis (66, 69, 109)	• Led to time savings (66) • Allows articles to be categorized (92)	• Demonstrates good sensitivity and specificity (66). Found to be useful when there is little manually-assigned information available (109)	• Manual set up can be time consuming (66) • Unable to analyse individual words – required "bags" of words (69)	Important to set up the technology with researchers to ensure they are confident in using it. Led to time savings (66)

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Automatic Classification	Classifies title and abstracts according as relevant or not. Researchers start by categorising articles as relevant and the tool uses those that are both included and excluded to build a model to screen the remaining articles (20, 40)	Screening, extraction, analysis (20, 40)	• Analyse large amounts of complex data across different subject areas (20) • Reduced manual screening and classification time while identifying 10x more relevant studies than manual approaches (20, 40)	—	• Technically difficult to use (20) • Large amount of computer power needed if large datasets are used (20) • Human input still needed to train tool and respond to information being produced (20) • Large number of articles meant process was still time consuming (20) • Results can be biased if initial sample is biased (40) • Can struggle to work with complex terms requiring greater interpretation (40)	Initial training needed to understand how to use the tool effectively (20) Challenges in identifying appropriate metrics to analyse the performance of the tool (20) Not sufficiently integrated into other review tools (40)
Automatic Term Recognition	Orders the list of articles for screening in order of relevancy and can extract terms from text (20, 40). The title and abstracts are reviewed first and a score applied to each article to indicate relevancy and a list of relevant terms are collated from the articles. The top 100 terms are used to search the unscreened articles to score these according to their relevancy. Researchers screen the top X and identify relevant articles and the process is re-run – this is repeated regularly with increasingly larger sets of articles (20)	Screening, extraction, analysis (20, 40, 44)	• Analyse large amounts of complex data across different subject areas (20) • Reduced manual screening time (estimates of 50%) while identifying 10x more relevant studies than manual approaches (20, 40). This allows full-text documents to be retrieved earlier and allows interim assessment (40)	• Technically easy to use (20)	• Large amount of computer power needed if large datasets are used (20) • Human input still needed to train tool and respond to information being produced (20) • Large number of articles meant process was still time consuming (20) • If initial sample is biased, results from using the tool may also be biased (40)	Initial training needed to understand how to use the tool effectively (20) Challenges in identifying appropriate metrics to analyse the performance of the tool (20)

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Bidirectional encoder representations from transformers (BERT)	A neural network that can understand contexts of sentences (22, 30)	Screening, extraction (22, 30, 46, 59)	—	• Advantage over other models as it is trained using English Wikipedia and BookCorpus so has been exposed to large amounts of text before it is trained for use in a specific review (30). There is also less of a need to pre-process data (59) • Can be used in combination with other machine learning algorithms and neural networks (30)	• Performance is lower than expected, possibly because a large enough data set was not used (30) • Human input and expertise needed to label data to train algorithm (30) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30)	—
DistillerSR	Machine learning and text mining approach based on Natural Language Processing used for title and abstract screening which can prioritise references to screen and allows custom screeners to be created (44, 121, 131, 159)	Screening, extraction, analysis (32, 44, 99, 121, 131, 159)	• Reduced screening burden for researchers (131)	• User friendly and easy to use (99) • Trusted (99)	• Best performance occurs in conjunction with manual screening; used alone, sensitivity is low and so cannot reliably be used alone (121) • One study showed no advantage of using the tool (in terms of time saving, articles missed and workload saved) compared to the baseline (99) • Greater reduction in manual workload associated with greater number of missed articles (99, 131) • Some tendency to over include articles (131) • Cannot distinguish between articles that were initially excluded and then re-included by another human reviewer (131) • Human input needed to clean data before the tool is used and develop a training set (131) • Accuracy of tool depends on quality of the training set (131)	Important to pilot the tool before conducting full screening (131)

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Naïve Bayes classifier	A machine learning algorithm used for text classification (30, 143)	Screening, extraction (30, 44, 46, 80, 134)	• Can reduce human workload (134)	• Demonstrated good effectiveness and speed compared to other text classification algorithms (30, 80) • Can be used in combination with other machine learning algorithms and neural networks (30, 80)	• Human input and expertise needed to label data to train algorithm (30, 134). Some forms of Naïve Bayes might not be accurate enough to classify abstracts alone (134) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30) • Variation in accuracy (80)	—
Support vector machine	A machine learning algorithm using linguistic analysis for text classification (30, 104, 138, 143)	Screening, extraction, analysis (30, 44, 52, 80, 104, 105, 115, 138)	• Find information faster by having annotations in articles (104) • Reduces the number of articles that need manual screening (105, 138) •	• Demonstrated good effectiveness compared to other text classification algorithms once it had been trained (30, 59, 104, 105, 115, 138) • Can be used in combination with other machine learning algorithms and neural networks (30, 80) • Some models do not need to rely on if the researchers knows ahead of time if topic specific training sets are available (115)	• Human input and expertise needed to label data to train algorithm (30) • Training set needs to be developed based on specific topic of focus (30) • Researchers discard articles from the training set classified as 'maybe' when this set should cover articles that are and are not relevant (30) • May not identify all articles, e.g. if they are peripherally relevant to the question (105, 138) • Tool does not perform as well if training set is skewed towards either article examples to include or exclude – need to ensure training set is balanced (138) • One study found the tool failed as a classifier, possibly due to violation of assumptions (52) • Variation in accuracy (80)	—
Machine learning based classifier tool (unnamed)	Automated citation classification system which can categorise articles as having high quality evidence or not. It is trained using manually coded exclusion criteria (127)	Screening, quality assessment, analysis (127)	• Fewer articles requiring manual review, saving researcher time (in some cases there can be more than 50% reduction in workload) (127)	• Demonstrates good performance (127) • Is quick and easy to use (127)	• Performance varied depending on research topic (127) • The decision-making process of the algorithm is unclear (127)	More work needed to refine classification systems and to determine which topics the algorithm works better on (127)

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Barzooka	Identified inappropriate use of bar charts for analysing continuous data (28)	Quality assessment (28)	—	—	—	—
CAMARADES meta-analysis	Automates meta-analysis and allows for visualisation of results, including to explore heterogeneity across studies (28)	Quality assessment, analysis (28)	—	—	—	—
Regular Expression Tool	Assess risk of bias in preclinical studies (28)	Quality assessment (28)	—	—	—	—
RobotReviewer	A web-based tool that automatically assess risk of bias (using distant supervision) and reviews sentences to determine if they are relevant to extract (22, 33, 41, 101, 107). It can produce a report to summarise key findings, e.g. research participants and intervention (107)	Quality assessment, extraction, analysis (21, 22, 32, 33, 41, 43, 44, 101, 107)	• Saves researcher time by not having to read full-texts (107)	• Can apply the Cochrane Risk of Bias tool (41, 101) • Similar (or better) accuracy to manual approaches, and performance is improving over time (41, 101, 107) • Freely available (101) • Use of distant supervision means algorithm can be trained on existing databases, rather than requiring human supervision (101)	• Requires human input to create training set for tool to function effectively (33, 41) • Could not extract some items (43) • Accuracy compared to human consensus can be low (101, 107). Quality assessment should be treated as a suggestion and should be manually reviewed and checked by researchers (33, 107) • Can only assess quality of articles as low or high/unclear which does not meet Cochrane guidance (101) • Only able to review a small number of sources of bias and cannot assess risk for more than one outcome per trial (101) • Can only be used on articles published online and in English (101) • Further research needed into use of tool in practice to assess efficacy (107)	—
SciScore	Identified the reporting of rigour criteria in articles (28)	Quality assessment (28)	—	—	—	—
ContentMine	Extracts structured data from tables and graphs embedded in PDFs (41)	Extraction (41)	—	—	—	—
ExaCT	Algorithm that labels sentences as relevant or not to the review (e.g. treatment frequency) and can distinguish control from intervention group (27, 33)	Extraction (27, 32, 33, 44)	—	—	• No association is made, e.g. between outcome with a trial arm (27) • The results are only available in HTML (27)	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Graph Digitization software	Algorithm using line recognition on graphs to help researchers trace plots and extract the raw data from graphs (27)	Extraction (27)	—	• Not used specifically for systematic reviews and so can bey used on common graphs (e.g. x-y and polar plots) (27)	• Cannot be used on survival curves, common in clinical trials (27)	—
Graph2Data	Extracts structured data from tables and graphs embedded in PDFs (28, 41)	Extraction (21, 28, 41)	—	—	—	—
Long short-term memory network		Extraction (46)	—	—	—	—
Name Entity Recognition	Similar to Natural Language Processing, it can be used to identify specific names (e.g. diseases, drugs) and numbers for meta-analysis (26)	Extraction (26)	—	—	—	—
NLP software (unnamed)	Natural language processing model used to extract PICO principles from unstructured text. The biomedical language variant of BERt was used as a basis for the model (152)	Extraction (152)	• Review unstructured data quickly to help inform public health decisions (152)	• Demonstrates good performance and a low error rate (152) • User friendly online platform (152)	• Needs some researcher input to put together the article dataset and check the accuracy of extracted information (152) • Performance is influenced by the accuracy of the training set (152)	—
R extraction tool (unnamed)	Extract specific sections of an article which can be used for PDFs (130)	Extraction (130)	—	—	• Did not work on a significant proportion of articles (130)	—
RelEx	Software that can extract numerical data, e.g. number of disease cases (136)	Extraction (136)	• Can be used for monitoring ongoing health events (136)	—	• Performance is influenced by accuracy of data being extracted, e.g. differences in language used to describe the same thing (136) • Challenge in grouping data together discussing the same event (136)	—
Rule-based approach to data extraction	Automatic extraction of data from systematic reviews (139)	Extraction (139)	• Reduce risk of producing redundant reviews (139)	• Demonstrated good efficacy at extracting relevant structured and semi-structured information which can be used to model the risk of conclusion change (139)	• Cannot be used for all literature databases (139) • The format of text can make extraction difficult (139)	More structured text in systematic reviews would benefit many automated tools in extracting data (139)
Sequence Tagging	Can model correlation between words using Natural Language Processing (22)	Extraction (22)	—	—	—	—

Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
TimeML	Algorithms that can extract temporal data from grey literature (136)	Extraction (136)	• Can be used for monitoring ongoing health events (136)	• Demonstrates performance nearly as good as manual extraction (136)	• Performance is influenced by accuracy of data being extracted, e.g. differences in language used to describe the same thing (136) • Challenge in grouping data together discussing the same event (136)	—
Automatic text summarisation	Provide a text or graphic summary of the most important themes to allow processing and interpretation of data more quickly(19, 145). It can include aspects of natural language processing and machine learning (19)	Extraction, analysis (19, 145)	• Can analyse information from across studies to integrate information with similar meanings and highlight contrasting information (19)	—	• Requires human input, e.g. to develop reference standards (19, 145) • Can only review data in English (19)	Further research needed on the efficacy/accuracy of the tool and to develop reference standards (19)
Lda2vec	Text mining approach to summarise key themes and trends in research (69)	Extraction and analysis (69)	—	—	• Unable to analyse individual words – required "bags" of words (69)	—
WebPlotDigitizer	Allows for the re-digitisation of data from graphs (27)	Extraction, analysis (27, 44)	—	—	• Cannot be used on survival curves (27) • Does not have optical character recognition (27)	—
Wordscores	An algorithm used for automated content analysis, an automated approach to extracting patterns from text using systematic coding based on word frequencies(34). It has been used predominantly in analysis of political and policy documents (34)	Extraction and analysis (34)	—	• Demonstrated good effectiveness and reliability compared to manual coding in extracting and comparing policy positions from text, including complex documents of varying length and produced by different authors (34) • Reported as being easy to use, flexible, simple and quick, analysing text in a few seconds (34) • Researchers do not need to understand the meaning of the text the algorithm is coding (34) • Can analyse text in any language (34)	• Human expertise is needed to inform the choice of reference text, requiring subject expertise (34) • Reduced accuracy when analysing shorter documents (34) • Some argue the analysis is simplistic and not nuanced enough as it relies on word frequencies, raising questions about the validity of results and applicability to more complex questions (34) • Use for review analysis may not capture all meaning so manual checks are required (34)	—
Aggregator	Machine learning model based on Medline metadata which collates articles written by the same author or part of the same trial (111)	Analysis (111)	—	• Demonstrates good accuracy in grouping articles (111)	• Restricted to clinical trials published in PubMed (111) • Can not identify plagiarism, the same publication in different journals or where authorship is deliberately obscured (111) • The format of text or missing information can cause issues (111)	Better training sets are needed (111)

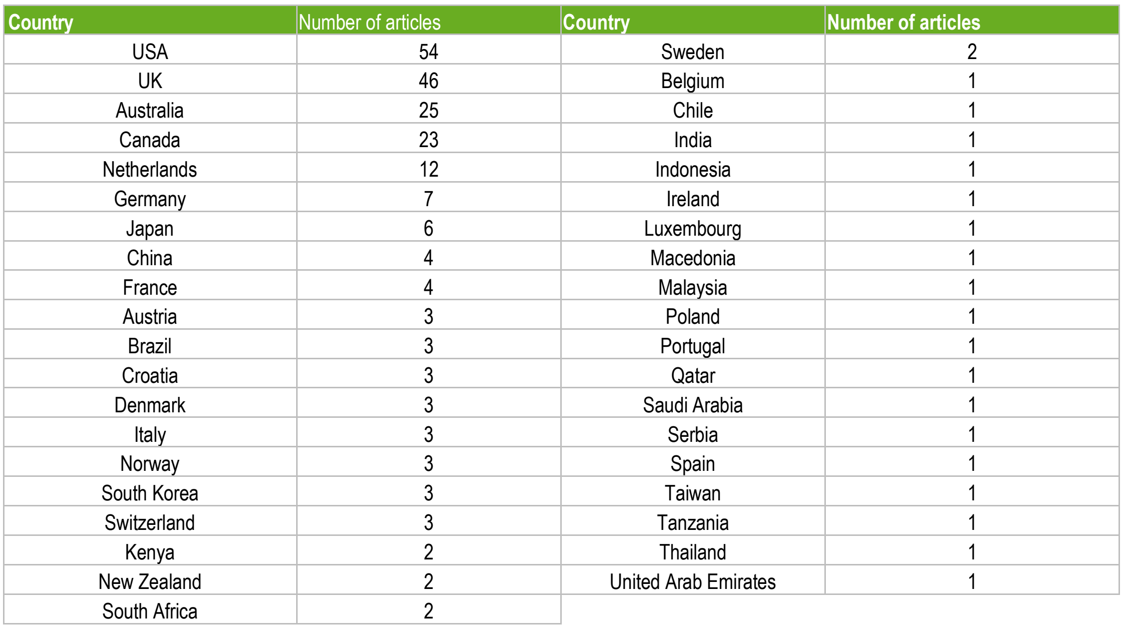
Type of technology	Description of technology	Stage	Impact of using the technology	What worked well	What did not work well	Gaps/needs in using the technology
Automated text analysis	Uses lexicon analysis to identify key words to summarise trends in research topics (160)	Analysis (160)	—	• Efficient and reliable tool (160)	—	—
Classification trees	Aspect of machine learning which produces a decision tree to model outcomes (82, 143)	Analysis (82)	• Effective at being able to estimate conclusion changes when updating systematic reviews (82) • Prevent unnecessary reviews being conducted (and funded) that do not provide new evidence which saves time (82)		• May not be able to be used for non-English language or non-clinical trial articles (82) • Only tested using small samples of articles (82)	Structured database of systematic reviews needed (82)
Hidden Markov Modelling	-	Analysis (44)	—	—	—	—
Meta-Analyst	Conducts meta-analysis on extracted data (27)	Analysis (27, 44)	—	—	• Limited ability to integrate with other extraction systems (27)	—
MetaPreg	—	Analysis (32)	—	—	—	—
MetaXL	—	Analysis (32)	—	—	—	—
Optical Character Recognition	—	Analysis (44)	—	—	—	—
Rule-based induction and machine learning (unnamed)	Lists and ranks evidence from biomedical literature (146)	Analysis (146)	—	• Demonstrated good performance (146)	• Some manual correction is needed (146)	—
Systematic EvidEnce Disseminator	-	Analysis (44)	—	—	—	—
Tesseract	-	Analysis (44)	—	—	—	—
Text Classifier	-	Analysis (44)	—	—	—	—
RevMan HAL	Automatically creates sections of a review based on predefined templates using quantitative data (41)	Analysis, reporting (27, 32, 41, 44)	• Allows tracing of quality ratings back to original article (21)		• Can only work with RevMan files (27)	—
PRISMA Flow Diagram Generator	Automatically generates PRISMA diagrams (27)	Analysis, writing-up (27, 44)	—	—	• Cannot produce complex diagrams (27)	—
Glossaryfication Web Service	Identified terms in articles and creates a list of matching definitions from online glossaries which can then be edited by researchers. This supports the interpretation of terms by the reader (161)	Writing-up (161)	• Prevents misinterpretation of results (161)	—	—	—
Natural Language Generation	Can automatically produce sections of reports, e.g. description of literature and summaries (27)	Writing-up (27)	—	—	• Can introduce errors in reporting, e.g. if format of data differs across literature (27)	—
GDT	—	Unspecified	—	—	—	—
GRADE Pro	—	Unspecified	—	—	—	—

Annex 2 Findings tables

Table 3. List of organisations contributing to the reviewed literature

Organisation	No. of articles
University of Manchester	10
Bond University; University College London	9
Northeastern University	7
Oregon Health & Science University; University of Alberta; University of Oxford; US National Library of Medicine	6
King's College London; University of London; University of Edinburgh; University of Ottawa	5
Toyota Technical Institute; Imperial College London; Macquarie University; Ottawa Hospital Research Institute; University College Cork; University of Bristol; University of Cambridge; University of New South Wales; University of Pittsburgh; University of Technology Sydney	4
Harvard Medical School; McGill University; Monash University; University of Notre Dame Australia; University of Nottingham; University of Toronto; University of Utah	3
American University; Danube University Krems; Greater Manchester Mental Health NHS Foundation Trust; Inserm; La Trobe University; Mayo Clinic; NIHR ARC West; NIHR Greater Manchester Patient Safety Translational Research Centre; Norwegian Institute of Public Health; Osaka University; Oxford Health NHS Foundation Trust; Public Health Agency of Canada; Public Health Wales NHS Trust; Purdue University; Sciome LLC; St. Michael's Hospital (UK); Swansea University; Tufts University; University of Amsterdam; University of California; University of Glasgow; University of Illinois; University of Liverpool; University of North Carolina; University of Split; University of Tasmania; University of Tokyo; University of Ulster; Virginia Mason Medical Center	2
Aarhus University; Agency for Healthcare Research and Quality (USA); Anthrophi Technologies; Arizona State University; Asia University; Asia University Hospital; Auckland University of Technology; Australian Institute of Health Innovation; Avenir Health; Beijing University of Posts and Telecommunications; Berlin Institute of Health; Boston Children's Hospital; Brandenburg Medical School Theodor Fontane; Brown University; Campus Universitaire; Canadian College of Naturopathic Medicine; Cardiff Metropolitan University; Catholic University of Croatia; Centers for Disease Control and Prevention; Central Michigan University; Centre National de Re´ fe´rence des Maladies Auto-Immunes Rares; Chiang Mai University;China Medical University Hospital; Chinese University of Hong Kong; Cochrane; Cochrane Australia; College of Information Technology; Concordia University; Cracow University of Economics; Curtin University; Dalian Dermatology Hospital; Dalla Lana School of Public Health; Data Republic; Deloitte Consulting; Doctor Evidence; ES-SO Valais-Wallis; Eulji University; Evidence-based Practice Center; Federal University of Sao Paulo; FRONTEO Healthcare Inc; Galgotias University; German Federal Institute for Risk Assessment; Getulio Vargas Foundation; GSK; Health Canada; Health Data Research UK; HEC Montréal; Herdecke University; Houston Community College; Hyogo College of Medicine; IDEAS Cente; Institute for Health Services and Health System Research (Germany) ; Institute of East West Medicine; Institute of Science and Technology (UK); Instituto de Investigación Biosanitaria; Intermountain Healthcare; International Center of Insect Physiology and Ecology; Iowa State University; Izumo Citizens Hospital; Johannes Gutenberg-University; Joint Research Centre, European Commission; Julius Kühn-Institut; Kansas University School of Medicine; Karolinska Institutet; Keele University; King Abdulaziz University; Konkuk University; Korea University Ansan Hospital; Korea University College of Medicine; Kyung Hee University; Kyungpook National University; La Jolla; Li Ka Shing Knowledge Institute; LIMSI-CNRS; London School of Economics and Political Science; London School of Hygiene and Tropical Medicine; Luxembourg Institute of Science and Technology; Maastricht University; Masinde Muliro University of Science and Technology; Massachusetts General Hospital; Max Planck Institute for Molecular Genetics; McMaster University; MD Anderson Cancer Centre; Mistra EviEM; Monash University Malaysia; MRC Integrative Epidemiology Unit; National Health Insurance Service Ilsan Hospital; National Institute for Health and Care Excellence; National Institutes of Health; National Library of Medicine; Nested Knowledge Inc; Netcompany A/S; North Carolina State University; Norwegian University of Science and Technology; Nottingham Trent University; Nutrition Research Australia; Ohio State University; Open Science Community Utrecht; Østfoldforskning AS; Ottawa Hospital; Pacific College of Health Sciences; Polish Academy of Sciences; Polytechnic Institute of Viseu; Portland VA Medical Center; Precision HEOR; Qatar university; Queen's University; Radboud University Medical Center; RAND Corporation; René Rachou Institute; Robert Gordon University; Roma Tre University; Ryerson University; Saints Cyril and Methodius University; Saitama Medical University; Sapienza University of Rome; Scuola Superiore Sant'Anna; Second Hospital of Tianjin Medical University; Semmes-Murphey Clinic; Seoul National University College of Medicine; Shimane University; Southern Cross University; St Luke's International University; St. Mary's Hospital Centre (UK); Stanford University; Sunshine Coast University Hospital; Superior Medical Experts Inc; Swinburne University of Technology; Taylors University; Technical University of Denmark; Technische Universität Dresden; The Australian e-Health Research Centre; The Institute of Statistical Mathematics; The Netherlands Organisation for applied scientific research; The Park Centre for Mental Health; The Second Affiliated Hospital of Dalian Medical University Dalian; The World Health Organization; Thoughtful Technology;Tianjin Institute of Cardiology; Tilburg University; Trip Database Ltd; Tsinghua University; Tufts Medical Center; Univeristy of Liverpool; Universidad de Valparaíso; Universidade da Beira Interior; Université de Genève; Université du Québec à Montréal; Université Paris Saclay; University Hospital of Giessen and Marburg; University Hospital Split; University Medical Center Utrecht; University of Aberdeen; University of Applied Sciences Western Switzerland; University of Auckland; University of Bath; University of Belgrade; University of Bern; University of British Columbia; University of Calgary;University of Cape Town; University of Central Florida; University of Connecticut; University of Copenhagen; University of Dar es Salaam-Mbeya; University of Florida; University of Freiburg; University of Geneva; University of Granada; University of Groningen; University of Indonesia; University of Iowa; University of KwaZulu-Natal; University of Melbourne; University of Munster; University of North Carolina School of Medicine; University of Novi Sad; University of Queensland; University of São Paulo; University of Science and Technology; University of Sheffield; University of Sydney; University of the Sunshine Coast; University of Texas; University of Washington; University Paris-SUD; Utrecht University; VA Medical Center; Vanderbilt University; VarMac Consulting Engineers; Vienna University of Technology; Vrije Universiteit Amsterdam; Wageningen Food Safety Research; Wageningen University & Research; Waikato Hospital; Waitemata District Health Board; Watson Health Cloud; Werribee Mercy Hospital; Xtract AI	1

Table 4. **Articles broken down by country**



Country	Number of articles	Country	Number of articles
USA	54	Sweden	2
UK	46	Belgium	1
Australia	25	Chile	1
Canada	23	India	1
Netherlands	12	Indonesia	1
Germany	7	Ireland	1
Japan	6	Luxembourg	1
China	4	Macedonia	1
France	4	Malaysia	1
Austria	3	Poland	1
Brazil	3	Portugal	1
Croatia	3	Qatar	1
Denmark	3	Saudi Arabia	1
Italy	3	Serbia	1
Norway	3	Spain	1
South Korea	3	Taiwan	1
Switzerland	3	Tanzania	1
Kenya	2	Thailand	1
New Zealand	2	United Arab Emirates	1
South Africa	2		

Annex 3 Full search strategy

Table 5. Search strategy 1: OVID Medline

Ovid MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-Review & Other Non-Indexed Citations and Daily 1946 to March 04, 2022 Search executed: 7 March 2022	Results
1 ((technolog* OR innovat* OR tool*) adj2 (new OR emerging OR innovat* OR digital OR novel OR advance*)).tw.	286145
2 automat*:ti,ab,kw OR 'semi automat*':ti,ab,kw OR 'machine assist*':ti,ab,kw OR 'artificial intelligence':ti,ab,kw OR ai:ti,ab,kw OR 'ai based':ti,ab,kw OR 'machine learning':ti,ab,kw OR 'machine learning-based':ti,ab,kw OR 'natural language processing':ti,ab,kw OR 'expert system*':ti,ab,kw OR 'neural network*':ti,ab,kw OR 'data mining':ti,ab,kw OR 'text mining':ti,ab,kw OR 'web crawl*':ti,ab,kw OR 'web scrap*':ti,ab,kw OR 'text classification*':ti,ab,kw OR crowdsourc*:ti,ab,kw OR 'crowd sourc*':ti,ab,kw OR 'citizen science':ti,ab,kw	566710
3 ((supervised OR unsupervised OR reinforcement) NEAR/2 (classification* OR learning OR cluster*)):ti,ab,kw	23759
4 'artificial intelligence'/de	38538
5 1 or 2 or 3 or 4	951497
6 ((review* OR synthes*s OR assessment*) NEAR/2 (literature OR evidence OR systematic OR scoping OR knowledge OR rapid OR expedit* OR living OR research)):ti	312052
7 'meta analys*':ti OR metaanalys*:ti OR 'meta research':ti	182291
8 'risk of bias':ti OR 'quality assessment':ti OR 'eligibility assessment':ti OR 'search string*':ti OR 'search strateg*':ti OR 'evidence mapping':ti	7759
9 ((Screen* OR select* OR retriev* OR identif* OR rank* OR extract*) NEAR/2 (article* OR literature OR evidence OR reference* OR title* OR abstract*)):ti	3492
10 6 or 7 or 8 or 9	416591
11 5 and 10	9303
12 #5 AND #10 AND ([article]/lim OR [article in press]/lim OR [review]/lim OR [preprint]/lim) AND [english]/lim AND ([embase]/lim OR [preprint]/lim) AND [2000-2022]/py	5154
removed duplicates with Medline	648

Table 6. Search strategy 2: Embase

EMBASE Search executed: 7 March 2022	Results
1 ((technolog* OR innovat* OR tool*) NEAR/2 (new OR emerging OR innovat* OR digital OR novel OR advance*)):ti,ab,kw	377630
2 automat*:ti,ab,kw OR 'semi automat*':ti,ab,kw OR 'machine assist*':ti,ab,kw OR 'artificial intelligence':ti,ab,kw OR ai:ti,ab,kw OR 'ai based':ti,ab,kw OR 'machine learning':ti,ab,kw OR 'machine learning-based':ti,ab,kw OR 'natural language processing':ti,ab,kw OR 'expert system*':ti,ab,kw OR 'neural network*':ti,ab,kw OR 'data mining':ti,ab,kw OR 'text mining':ti,ab,kw OR 'web crawl*':ti,ab,kw OR 'web scrap*':ti,ab,kw OR 'text classification*':ti,ab,kw OR crowdsourc*:ti,ab,kw OR 'crowd sourc*':ti,ab,kw OR 'citizen science':ti,ab,kw	566710
3 ((supervised OR unsupervised OR reinforcement) NEAR/2 (classification* OR learning OR cluster*)):ti,ab,kw	23759
4 'artificial intelligence'/de	38538
5 1 or 2 or 3 or 4	951497
6 ((review* OR synthes*s OR assessment*) NEAR/2 (literature OR evidence OR systematic OR scoping OR knowledge OR rapid OR expedit* OR living OR research)):ti	312052
7 'meta analys*':ti OR metaanalys*:ti OR 'meta research':ti	182291
8 'risk of bias':ti OR 'quality assessment':ti OR 'eligibility assessment':ti OR 'search string*':ti OR 'search strateg*':ti OR 'evidence mapping':ti	7759
9 ((Screen* OR select* OR retriev* OR identif* OR rank* OR extract*) NEAR/2 (article* OR literature OR evidence OR reference* OR title* OR abstract*)):ti	3492
10 6 or 7 or 8 or 9	416591
11 5 and 10	9303
12 #5 AND #10 AND ([article]/lim OR [article in press]/lim OR [review]/lim OR [preprint]/lim) AND [english]/lim AND ([embase]/lim OR [preprint]/lim) AND [2000-2022]/py	5154
removed duplicates with Medline	648

Table 7. Search strategy 3: Cochrane

COCHRANE [Cochrane.com – via Wiley] Search excuted 7 March 2022	Results
1 ((technolog* OR innovat* OR tool*) NEAR/2 (new OR emerging OR innovat* OR digital OR novel OR advance*)):ti,ab,kw	15565
2 automat*:ti,ab,kw OR 'semi automat*':ti,ab,kw OR 'machine assist*':ti,ab,kw OR 'artificial intelligence':ti,ab,kw OR ai:ti,ab,kw OR 'ai based':ti,ab,kw OR 'machine learning':ti,ab,kw OR 'machine learning-based':ti,ab,kw OR 'natural language processing':ti,ab,kw OR 'expert system*':ti,ab,kw OR 'neural network*':ti,ab,kw OR 'data mining':ti,ab,kw OR 'text mining':ti,ab,kw OR 'web crawl*':ti,ab,kw OR 'web scrap*':ti,ab,kw OR 'text classification*':ti,ab,kw OR crowdsourc*:ti,ab,kw OR 'crowd sourc*':ti,ab,kw OR 'citizen science':ti,ab,kw	30496
3 ((supervised OR unsupervised OR reinforcement) NEAR/2 (classification* OR learning OR cluster*)):ti,ab,kw	453
4 [mh "artificial intelligence"]	1314
5 1 or 2 or 3 or 4	45964
6 ((review* OR synthes*s OR assessment*) NEAR/2 (literature OR evidence OR systematic OR scoping OR knowledge OR rapid OR expedit* OR living OR research)):ti	4292
7 'meta analys*':ti OR metaanalys*:ti OR 'meta research':ti	6867
8 'risk of bias':ti OR 'quality assessment':ti OR 'eligibility assessment':ti OR 'search string*':ti OR 'search strateg*':ti OR 'evidence mapping':ti	3517
9 ((Screen* OR select* OR retriev* OR identif* OR rank* OR extract*) NEAR/2 (article* OR literature OR evidence OR reference* OR title* OR abstract*)):ti	100
10 6 or 7 or 8 or 9	13472
11 5 and 10	546
12 not records CT.gov & ICTRP	481
13 **removed internal duplicates** removed duplicates with medline and embase	393

Annex 4 Further methodology detail

Further detail on study selection

Table 8. Inclusion and exclusion criteria

	Inclusion	Exclusion
Topic	Evidence syntheses for any subject	N/A
	The actual availability or use of any automated technology involving an element of automation and/or learning, used for evidence syntheses	Any digital technologies that considered do not include any element of automation or learning.
	What automated technologies have been used, how and by whom.	The use of automated technologies for any other purposes
Study type	Reviews, systematic reviews, meta-analyses, commentaries, editorials	Empirical studies, theoretical papers, conference proceedings that do not include full text, letters, review protocols
Date	Published since 2000	N/A
Language	All languages	N/A

Data extraction templates

The extraction templates for each type of extraction were similar, with the light extraction being more simplistic (see Chapter 2 for further details).

Table 9: Full extraction template

Basic information				Technology			
Citation	Document or study type	Short description of document	Topic or subject area	Name of technology	Description of technology	Existing technology or developed by authors?	Other information on technology

Use of technology in evidence synthesis						Experience using technology				
Reason for using technology	Stage(s) of evidence synthesis	Positive impacts of using technology	Negative impacts of using technology	How might the technology be used in the future?	Other information on use of technology in evidence synthesis	What was needed to use technology effectively?	What worked well?	What did not work well?	Gaps/needs in using the technology	Other information on experiences and lessons learned

Mapping						Other/reflections	
Authors organisations	Location of organisation	Relevant contacts (e.g. authors) and contact information where available	Infectious disease relevant (Y/N)	COVID relevant (Y/N)	Other information relevant to mapping	Researcher reflections	Other relevant sources or contacts

Table 10: Light extraction template

Basic information				Technology						
Citation	Document or study type	Short description of document	Topic or subject area	Name of technology	Description of technology	Stage(s) of evidence synthesis	What worked well?	What did not work well?	Gaps/needs in using the technology	Other information on technology

Mapping						Other/reflections	
Authors organisations	Location of organisation	Relevant contacts (e.g. authors) and contact information where available	Infectious disease relevant (Y/N)	COVID relevant (Y/N)	Other information relevant to mapping	Researcher reflections	Other relevant sources or contacts

Interview and focus group analysis coding frame

Table 11: Qualitative data coding frame

Node	Sub-nodes
Current approaches to evidence syntheses	How are reviews conducted and experience of this
	Has approach to reviews changed during COVID-19
	Other
Challenges of traditional evidence synthesis	Burden or cost of conducting reviews
	Duplication or redundancy
	Interpretation of evidence
	Time to write and publish
	Volume of literature
	Other
Advantages of automation	Collaboration
	Enables completion of new tasks
	Identification of more relevant literature
	Reduce redundancy
	Reduction in burden and costs of reviews
	User friendly
	Other
Challenges of automation	Continued need for human input
	Cost
	Equipment or research requirements
	Lack of guidelines or standards for automation
	Need for new staff training
	Technical limitations
	Other
Gaps or needs in using automation	Collaborations, networks or partnerships
	Resources
	Skills, knowledge or experience
	Other
Support in using automation	Collaborations, networks or partnerships
	Resources
	Skills, knowledge or experience
	Other
Software suites	Covidence
	DistillerSR
	EPPI Reviewer
	Rayyan
	Other

Annex 5 Focus group protocol

Introductions

1. Could you briefly go around the virtual room and introduce yourselves, including your name, organisation and a brief description of your experience of conducting evidence reviews (including the use of automation, if applicable)?

Approaches to evidence synthesis (general)

2. How do you or your organisation usually conduct evidence syntheses? *E.g. to keep abreast with scientific evidence, to inform public health decision-making.*
    a. What type of tools do you use (automated or otherwise)?
    b. What is your overall experience of conducting evidence reviews?
    c. Did the way you conduct evidence reviews change during the pandemic? If so, how? If not, why not?

Use and experiences of automated technologies for evidence synthesis

3. What are some of the biggest challenges you face in keeping up to date with evidence in your field? *E.g. burden/cost of conducting reviews, volume of literature, interpretation of evidence, duplication/redundancy.*
    a. Can automating or semi-automating parts of the evidence synthesis process help overcome some of these challenges? Why or why not?
    b. At which point(s) in the evidence synthesis process could (or has) automation be useful *(designing research questions, developing search protocol; running literature searches; screening articles; data extraction; risk of bias assessments; extracting data; analysis, reporting)*?
4. Have you (or others in your organisation) ever automated any aspect of an evidence review?
    a. If yes, why? *E.g. reduce burden/cost of conducting reviews, identify more (relevant) literature, improve interpretation of evidence, enable you to do new tasks, influence decision/policy making, develop/update guidelines or recommendations, reduce duplication/redundancy, ability to coordinate with ECDC, monitor the pandemic?*
    b. If not, why not?
    c. How did you determine which tool(s) were best for your needs?
    d. How would you describe the experience of using automation?
        i. What works well or supports the use of automated technology for evidence syntheses? *E.g. good performance/efficacy/validity of tool, user-friendliness, ease of use, useful features, having experienced/trained staff, having appropriate budget or equipment, having guidance on robust use of automation*
        ii. What are some of the sticking points in terms of being able automate part of the evidence review process? *E.g. technical limitations of tool, continued need for (significant) human input, challenges in accessing technology, need for new staff or training, resource/equipment requirements, difficult getting journal acceptance, lack of guidelines/standards for using automation, accessing technology.*
    e. For those who have used automated technologies in the past, would you have any advice for other organisations that might use automation in the future?

Needs and gaps in using technology for evidence synthesis

5. Thinking about the challenges we discussed earlier in the use of automation, how can these be overcome?
    a. What is needed to be able to use automation effectively? *E.g. skills/knowledge, experience/certain roles/staff, financial resources, equipment, facilities, partnerships/networks, technical aspects, new or updated guidelines/standards needed on the robust use of automation, changes to the way literature is currently reported, stored and indexed*
        i. Who would need to provide these?

    b. Is there anything else that would support your use of automation for evidence synthesis?

6. Who might it be useful to collaborate with to help fill gaps and address challenges in the use of automation?
    a. For those who have used automation, did you collaborate with or learn from other organisations when you began to use this technology?
    a. What would you hope to learn from other organisations?
    b. Would it be useful to collaborate with other public health competent authorities? Which ones?
    c. Is there support that ECDC could provide to coordinate and support the use of technology in public health?

Closing and follow-up

7. Is there anything else that you feel would be relevant that hasn't already been covered?
8. In the chat, we will share a link to a 2 minute follow-up survey to collect your feedback on today's session, including an option for you to express interest in being involved in further discussions on this topic. We would appreciate it if you could complete this survey soon after today's session.
    a. Would there be interest among the group to continue these discussions and collaborations?

Annex 6 Interview protocol

Introductions

1. Could you briefly say a bit about your role and of your experience of conducting evidence reviews (including the use of automation, if applicable)?

Approaches to evidence synthesis (general)

2. How do you or your organisation usually conduct evidence syntheses? *E.g. to keep abreast with scientific evidence, to inform public health decision-making.*

    a. What type of tools do you use (automated or otherwise)?

    b. What is your overall experience of conducting evidence reviews?

    c. Did the way you conduct evidence reviews change during the pandemic?

        i. If so, how? If not, why not?

Use and experiences of automated technologies for evidence synthesis

3. What are some of the biggest challenges you face in keeping up to date with evidence in your field? *E.g. burden/cost of conducting reviews, volume of literature, interpretation of evidence, duplication/redundancy.*

    a. Can automating or semi-automating parts of the evidence synthesis process help overcome some of these challenges? Why or why not?

    b. At which point(s) in the evidence synthesis process could (or has) automation be useful *(designing research questions, developing search protocol; running literature searches; screening articles; data extraction; risk of bias assessments; extracting data; analysis, reporting)*?

4. Have you (or others in your organisation) ever automated any aspect of an evidence review?

    a. If yes, why? *E.g. reduce burden/cost of conducting reviews, identify more (relevant) literature, improve interpretation of evidence, enable you to do new tasks, influence decision/policy making, develop/update guidelines or recommendations, reduce duplication/redundancy, ability to coordinate with ECDC, monitor the pandemic?*

    b. If not, why not?

    c. How would you describe the experience of using automation?

        i. What works well or supports the use of automated technology for evidence syntheses? *E.g. good performance/efficacy/validity of tool, user-friendliness, ease of use, useful features, having experienced/trained staff, having appropriate budget or equipment, having guidance on robust use of automation*

        ii. What are some of the sticking points in terms of being able automate part of the evidence review process? *E.g. technical limitations of tool, continued need for (significant) human input, challenges in accessing technology, need for new staff or training, resource/equipment requirements, difficult getting journal acceptance, lack of guidelines/standards for using automation, accessing technology.*

    d. *If interviewee has used technology:* Would you have any advice for other organisations that might use automation in the future?

Needs and gaps in using technology for evidence synthesis

5. Thinking about the challenges we discussed earlier in the use of automation, how can these be overcome?

    a. What is needed to be able to use automation effectively? *E.g. skills/knowledge, experience/certain roles, financial resources, equipment, facilities, partnerships/networks, technical aspects*

        i. Who would need to provide these?

6. Who might it be useful to collaborate with to help fill gaps and address challenges in the use of automation?

a. *If interviewee has used technology:* Did you collaborate with or learn from other organisations when you began to use this technology?

d. What would you hope to learn from other organisations?

e. Would it be useful to collaborate with other public health competent authorities? Which ones?

f. *For ECDC authorities:* Is there support that ECDC could provide to coordinate and support the use of technology in public health?

Closing and follow-up

7. Is there anything else that you feel would be relevant that hasn't already been covered?
8. Would you be interested in being involved in further discussions and collaborations on evidence reviews and the use of technology?

Annex 7 Survey protocol

Demographic information

1. What country do you primarily work in? [Drop down with EU/EEA countries, and 'other, please specify'.
2. What is the name of your organisation? [free-text box]
3. Which of the following options apply to you in your use and conduct of evidence syntheses/literature reviews? Please select all that apply:
   - I use evidence syntheses/literature reviews produced by others.
   - I conduct evidence syntheses/literature reviews.
   - None of the above. I do not use or conduct evidence syntheses/literature reviews.

Follow up for those that do not use or conduct evidence syntheses/literature reviews

4. Why don't you conduct or use evidence syntheses/literature reviews? [free-text box]
5. What would help you use evidence syntheses/literature reviews produced by others, and conduct evidence syntheses/literature reviews for yourself? [free-text box]

Using evidence syntheses/literature reviews

6. To what extent do you agree with the following statements? [rate as strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree or NA or I don't know].
   - It is difficult to keep up to date on public health topics that are important to my organisation due to the large volume of literature that is published in this area.
   - It is difficult to keep up to date on public health topics that are important to my organisation due to the delays in publishing, meaning that by the time evidence is published, it is already out of date.
   - It is difficult to keep up to date on public health topics that are important to my organisation due to a lack of evidence synthesis- and literature review-related skills and knowledge (lack of capability).
   - It is difficult to keep up to date on public health topics that are important to my organisation due to other competing priorities (e.g. other functions that I need to do in my professional role, creating a lack of capacity).
   - It is difficult to keep up to date on public health topics that are important to my organisation due to a lack of resources within my organisation (e.g. staff, funding, time).
7. Have you used evidence syntheses/literature reviews in the area of public health that have been conducted using automated or semi-automated technologies? [Yes/no/I don't know]
8. To what extent do you agree with the following statement? [rate as strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree or NA or I don't know].
   - I trust information that comes from evidence syntheses/literature reviews that use technologies to automate or semi-automate parts of the evidence synthesis/literature review process.
9. Are there other challenges not covered above that make it difficult to keep up to date on the latest evidence in the area of public health? If so, please provide more information. [free-text box]
10. What would help you more effectively use information from evidence syntheses/literature reviews to inform your work in public health? [free-text box]

Conducting evidence syntheses/literature reviews

11. Have you used automated or semi-automated technologies to assist in conducting evidence syntheses/literature reviews? This might be, for example, machine learning, natural language processing, crowdsourcing, text mining, neural networks, or any other tool/software that is used to automatically complete an aspect of an evidence synthesis/literature reviews task. [Yes/no]

Conducting evidence syntheses/literature reviews using automated technologies

12. Please list what technologies you have used in each step of the evidence synthesis/literature review

process. If you have not used technology for a particular step, please leave blank.

- Developing research questions/protocol
- Developing search strategy
- Searching for literature
- Screening literature (title and abstract)
- Screening literature (full text)
- Extracting information and data from literature
- Risk of bias assessment
- Analysing and synthesising results
- Writing up results
- Other parts of evidence synthesis/literature review process (please specify)

13. How did you determine which tool(s) best suit your needs? [free-text box]
14. What are the key challenges that you faced in using automated or semi-automated technologies in conducting evidence syntheses/literature reviews? Please select all that apply.
    - Lack of information on what technology would be best suited to my needs and/or what technologies are available to me
    - Lack of skills or knowledge in using evidence synthesis/literature review technology within organisation
    - Lack of resources within organisation (e.g. financial, equipment/facilities)
    - Lack of time/capacity to implement and learn how to use a new technologies
    - Difficulty adopting new technologies during public health emergencies
    - Concern around transparency, robustness and/or performance of automated and semi-automated methods
    - Lack of interoperability between technologies and/or a lack of a technology that can be used from start to end of the evidence synthesis/literature review pathway
    - Inability to collaborate with and learn from other organisations that have used similar technologies for evidence syntheses/literature reviews
    - My organisation did not face any of these challenges
    - Other (please specify)
15. How were these challenges overcome, if at all? [free-text box]

Conducting evidence syntheses/literature reviews without automated technologies

16. Why have you not used automated or semi-automated technologies to help with the evidence synthesis/literature review process yet?
    - Lack of information on what technology would be best suited to my needs and/or what technologies are available to me
    - Lack of skills or knowledge in using evidence synthesis/literature review technology within organisation
    - Lack of resources within organisation (e.g. financial, equipment/facilities)
    - Lack of time/capacity to implement and learn how to use a new technologies
    - Difficulty adopting new technologies during public health emergencies
    - Concern around transparency, robustness and/or performance of automated and semi-automated methods
    - Lack of interoperability between technologies and/or a lack of a technology that can be used from start to end of the evidence synthesis/literature review pathway

- Inability to collaborate with and learn from other organisations that have used similar technologies for evidence syntheses/literature reviews
- We have not considered using technologies to automate or semi-automate parts of the evidence synthesis/literature review process
- My organisation did not face any of these challenges
- Other (please specify)

17. What would be helpful in enabling you to use technologies to automate or semi-automate parts of the evidence synthesis/literature review process? [free-text box]
18. Which organisation(s), networks or groups, if any, would you find it useful to collaborate with to support the use of technology for conducting evidence syntheses/literature reviews? [free-text box]

Closing and follow up

19. Would you be interested in hearing about future discussions, events and collaborations around the use of technology in evidence syntheses/literature reviews? **[Yes/no]**

Contact details for follow up

20. Please provide your name and contact details below.