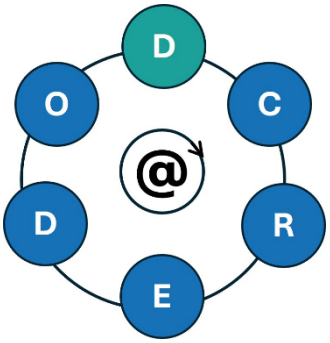
Co-funded by the European Union

D-CREDO

Digital Health Technologies-Augmented Clinical Reasoning Education



2024-1-PL01-KA220-HED-000247790

D2.1 Literature review report

Deliverable number	D.2.1
Delivery date	November 30, 2024
Status	Draft
Authors	Jonas Verdonschot (EMC), Laura Zwaan (EMC), Andrzej Kononowicz (JU), Charlotte van Sassen (EMC)

Document Revision Table

Version	Date	Author(s)	Organization	Comments
0.1	25.11.24	Jonas Verdonschot	EMC	First draft prepared.
0.2	26.11.24	Andrzej Kononowicz	JU	Updated draft, added structure.
0.3	27.11.24	Jonas Verdonschot	EMC	Updated draft, process feedback, added results.
0.4	27.11.24	Laura Zwaan	EMC	Update and edit draft.
0.5	29.11.24	Jonas Verdonschot	EMC	Process feedback, edit draft.
1.0	30.11.24	Jonas Verdonschot	EMC	Finalizing report.

Co-funded by
the European Union

Summary

Objectives: The document presents a report on the D-CREDO activity to produce an inventory of relevant literature, which will inform the development of learning objectives and the conceptual framework for the educational model, and, in the longer term, the project's learning units and guidelines.

Approach: A rapid review was designed and implemented to identify a set of relevant papers published in the last 5 years. A literature search was developed to include keywords related to clinical reasoning, the five types of D-CREDO digital tools, and the educational sciences domain. Five databases were searched: Medline (via Ovid), Embase, Web of Science, PsycInfo, and ERIC. The review process was managed using the Covidence management system. The consortium developed eligibility criteria structured by the PICOS framework and a data extraction form to capture data from the studies. Each step of the study selection process was performed independently by two reviewers, with any conflicts resolved by a third reviewer.

Results: The implementation of the search strategy resulted in 2,876 records, which, after the removal of duplicates, yielded a final dataset of 1,527. Abstract screening excluded 1,420 irrelevant studies. A total of 103 studies were retrieved for full-text analysis and description. Following this analysis, 57 studies were excluded for various reasons, with descriptions of the exclusion reasons provided. The remaining 46 studies were data-extracted and are available as a spreadsheet for reference in other project activities. The data show a growing and diverse body of research focused on the integration of digital health technologies in clinical reasoning education. While there is strong emphasis on Large Language Models (LLMs), electronic health records (EHRs), and learning activities, there is a noticeable gap in research related to education of clinical reasoning in the context of telehealth, mobile apps, and artificial intelligence (AI) in image analysis. Most studies are conducted in high-income countries. The range of study designs highlights the varied approaches being used, but there remains an underrepresentation of lower-income countries and certain educational and theoretical frameworks. This expanding body of research underscores the increasing importance of digital tools in clinical reasoning education, though further exploration in some areas is needed.

Conclusion: The first completed task under WP2 was the development of a literature database to inform the educational framework. It identified several helpful studies but also revealed that some of the D-CREDO tools were not well represented in the identified literature. Moving forward, the data gathered from this review will inform the development of the D-CREDO curriculum.

Co-funded by
the European Union

Table of Contents

1. Introduction and objectives

1.1. Motivation

The primary goal of Work Package 2 (WP2) is to develop a solid conceptual and theoretical basis for the student and train the trainer learning units planned within the D-CREDO project work package 3. As our plan is to build our curriculum on evidence, we emphasise the importance of incorporating in the educational model findings from recent literature.

Even though the partners of D-CREDO analysed the literature in preparation for the D-CREDO proposal, the progress in the field of digital health technologies is immense. The rapid integration of digital health tools into clinical practice presents new opportunities and challenges for medical and nursing education. Working as a consortium we have a unique opportunity to share the workload in a larger group of researchers to efficiently review in a systematic way a larger body of literature to inform the development of an educational framework for the D-CREDO curriculum. For that reason we have decided to incorporate in the agenda a rapid review activity.

The rapid review method [Gough17], [Tricco17] allows researchers to pose questions to literature queries that are answered in an agile process in a limited amount of time to address operational tasks as needed for instance in developmental projects. Planning the activity to last three months, as is common for rapid reviews, we have deliberately set constraints on the comprehensiveness and also types of analyses performed. This approach ensures speed and aligns with the 3 months time constraint to update our knowledge base on digital tools used in clinical reasoning (CR). Our overarching goal was to highlight key publications that will inform the development of the learning objectives (deliverable D2.2 due in M6) and conceptual framework (deliverable D2.3 in M9). With that in scope we work in the D2.1 on forming a project's "bookshelf" (a repository of recommended helpful studies) that can be referred to at the learning unit development and further extended with new literature updates when needed.

1.2. Research Goal

To focus the attention in this activity we have developed the following research question to guide us in the review:

How can digital health tools be utilized to enhance clinical reasoning education. Specifically, how should the use of digital health tools be taught and integrated into clinical reasoning practice and education for undergraduate medical and nursing students?

This deliverable report outlines the process and outcome of the rapid review aimed at providing evidence-based insights into teaching strategies, learning objectives, and methodologies for equipping future healthcare professionals with the skills to effectively use digital tools in clinical reasoning education.

1.3. Quality Criteria

In alignment with the the key performance indicators defined at the stage of writing the D-CREDO proposal, we set the following specific performance indicators to be addressed in the activity

- search query with selection of relevant keywords that include ≥1 for each D-CREDO type of digital tools
- ≥1 reviewer from each partner institution, representing educational, health, and technical background
- use of an online review management tool
- analysis of papers published since 2019
- inclusion/exclusion decision on ≥1000 reviewed abstracts
- ≥100 commented included papers (≥20 each partner)
- inclusion of studies describing educational theories, learning and assessment methods, digital tools in the context of CR practice and education at various stages of the process and ≥2 health professions
- commented set of selected relevant papers

2. Methods

The rapid review was carried out by a WP2 rapid review working group coordinated by EMC. The premise was that each project partner will designate at least one representative to participate in the working group from each D-CREDO partner institution. The inquiry yielded a far better result than anticipated with most partners proposing three group members to participate in the activity. The groups met regularly in dedicated “alignment meetings” on Zoom and Teams.

2.1. Conceptualisation of the search strategy

The first step in the review was the conceptualization of the search strategy. For that a working document was prepared in close collaboration between EMC (work package lead) and the JU (project coordinator) that resulted in the list of working definitions of the key concepts and keywords. This included three basic facets (aspects) to be included in the literature search:

- Aspect #1: Domain of clinical reasoning
- Aspect #2: D-CREDO Digital Tools
- Aspect #3: Educational aspect

A well-known challenge in the literature is the highly heterogeneous terminology used to describe clinical reasoning, with little consensus on definitions, even for core concepts [Young19],[Huseman23]. To address this challenge, the first step in developing an effective search strategy was to identify adequate keywords. Drawing on insights from previous research in the field, we compiled a list of terms frequently referenced in studies on the topic to describe the domain of clinical reasoning (aspect #1). This list included combinations of keywords such as *Clinical, Diagnostic, Therapeutic, Management* and *Analytic* with terms like *Reasoning, Thinking, Decision-making, Problem-solving,* and *Judgment*.

The second aspect regarded the five categories of D-CREDO tools that were elected in the process of general needs analysis prior to the project. This including the following five tools:

- AI in Image Analysis

- LLM (Large Language Models) & Big Data
- mHealth Apps & Wearables
- EHR (Electronic Health Records) & CDSS (Clinical Decision Support Systems)
- Telehealth

AI in image analysis refers to the use of AI-based tools for automatic image interpretation, encompassing opportunities such as mass-scale screening and challenges like dataset biases, automation bias, and the limited transferability of models across countries. These tools support clinical reasoning through tasks such as segmentation and highlighting regions of interest in case-based, AI-assisted image classification, with a special emphasis in fields like radiology.

The suggested keywords included: *Artificial intelligence, Deep Learning, Convolutional Neural Network, Computer Vision, Radiomics*

Large Language Models and Big Data. Large language models such as GPT-4, is the newest generation of AI tools that gained substantial interest after November 2022 due to its revolutionary capabilities in generation of natural language content. In health professionals' LLM support tasks like translation, generating knowledge summaries, identifying rare diseases, and AI-assisted data transformation. Big data analytics involve the oversight and utilisation of data from sources like social media to provide lifestyle and health improvement advice, with associated risks related to privacy, digital professionalism, and ethical considerations in the context of CR.

The suggested keyword included: *Large Language Model, Generative AI, Natural Language Processing, Prompt Engineering, ChatGPT, Big data*

mHealth apps & wearables are devices (e.g., smartwatches, bands, textiles) and software for mobile devices like smartphones that facilitate real-time and distributed data collection and activity tracking, and condition monitoring, influencing person-centred aspects of clinical reasoning. They are used as mental well-being apps in mental health, as a remote data collection in austere conditions, such as in war zones and providing accessible and adaptable healthcare solutions.

The suggested keyword included: *Smartphone, Digital sensor, Digital phenotyping, Health App, Mobile Health*

EHR and CDSS. Electronic Health Records (EHR) are systems used for medical documentation that include elements relevant to clinical reasoning, such as patient histories, observational notes, discharge letters, and lab values. However, they also present some challenges in the domain of clinical reasoning, including the need to gather information scattered across multiple patient notes, inconsistencies, copy-and-paste errors, difficult-to-decipher acronyms, and information overload. Clinical Decision Support Systems (CDSS) are tools integrated into healthcare processes and software, including electronic health records. These tools assist in generating differential diagnoses based on list of symptoms, monitoring actions, checking entered value ranges, suggesting tasks to perform, and issuing reminders.

The suggested keyword included: *Electronic Health Record, EMR, symptom checker, differential diagnosis generator, Clinical Decision Support Systems*

Finally, by **Telehealth** we meant the use of digital communication technologies to provide healthcare services remotely. It includes real-time virtual visits, where patients interact with healthcare providers through video or audio calls for diagnosis and treatment. It also covers asynchronous telemedicine, where medical information is transmitted between healthcare providers or between patients and providers at different times. In the context of clinical reasoning decision making around virtual visits for diagnosis and treatment and asynchronous sharing of medical information between healthcare providers. It also includes tasks related to telemonitoring and telerehabilitation.

The suggested keyword included: *Telehealth, Virtual visit, Virtual care, Telemedicine, Remote Patient Monitoring*

The third aspect of the search strategy aimed to include studies describing educational theories, learning and assessment methods. The professions of choice according to the project proposal were medical and nursing students. The search specifically included studies focusing on undergraduate medical and nursing students, explicitly excluding postgraduates and similar levels, to align with the scope of the D-CREDO Project. This subject heading and free-text keywords for this aspect covered e.g. as *nursing/medical education, undergraduate education, learning objectives, curriculum development, teaching process*.

To ensure the focus remained on the core objective of facilitating clinical reasoning education, it was critical that all three aspects—clinical reasoning, digital tools, and educational relevance—were present in the studies selected. Tools that merely addressed logistical aspects of organizing or delivering education, such as scheduling systems or communication platforms, were not sufficient for inclusion.

With the list of established requirements regarding the search strategy the deliverable coordinator from EMC contacted the medical library at Erasmus University to develop in collaboration with a professional information scientist experienced in systematic reviews a professional search strategy.

2.2. Selection of databases

In rapid reviews the number of searched databases is usually limited to 2-3 to fit the time constraints. However, we have been advised by the information specialist from Erasmus University Medical Library to include five databases for a high quality review result. This covered:

- **Medline ALL** (via Ovid): a standard reference in biomedical research maintained by the US-based National Library of Medicine
- **Embase** (via Embase.com): The “European Medline” maintained by Elsevier published known for a the best coverage of European journals and also speciality in publications on medical devices
- **Web of Science Core Collection** (via Web of Knowledge): Covers high quality research papers without limitation to the biomedical disciplines which also includes a broad range of disciplines including also studies published in technical, educational and social sciences.
- **PsycINFO** (via Ovid): As clinical reasoning deals with aspects of cognitive psychology we felt it is also crucial to cover database indexing research in psychology
- **ERIC** (via Ovid): Which is a leading database for education-related topics.

The search strategies were designed and piloted on a set of preselected studies of high quality in an iterative process between the WP2 working group leaders, project coordinator and the librarian. The final

detailed search strategies in selected databases, including the full list of search terms, are provided in Appendix 1.

2.3. The review process

The WP2 team comprised 13 reviewers from academic institutions across Europe, offering a diverse range of expertise in medical education and digital health. This interdisciplinary group included professionals from specialized fields such as healthcare education, clinical practice, clinical reasoning, pharmacology, medical physics, information technology, and bioengineering. Many team members held advanced qualifications, including PhDs and professional doctorates, which significantly enhanced the expertise contributing to the study. The team included experts affiliated with renowned institutions such as Bukovinian State Medical University (Ukraine), Erasmus MC (Netherlands), Jagiellonian University (Poland), UMIT Tirol (Austria), and Instruct (Germany). This blend of academic and professional knowledge, combined with regional and institutional perspectives, provided a rich, interdisciplinary approach to the study.

To maintain high standards of objectivity, each step of the study selection process was conducted independently by two reviewers. Any conflicts were resolved by a third reviewer affiliated with Erasmus MC. This process ensured consistency, minimized bias, and strengthened the reliability of the findings. Throughout the review process, the rapid review team held regular meetings coordinated and led by Erasmus MC. These meetings provided clear instructions and created opportunities for reviewers to ask questions, fostering alignment among participants. Three dedicated alignment meetings specifically focused on the rapid review, complemented by regular full-group meetings where the review was also discussed, as well as one-on-one discussions with team members.

These meetings were critical for maintaining consistency and addressing uncertainties during the review. Detailed emails containing all relevant information and instructions were shared, and reviewers were encouraged to reread these as needed. Smaller group discussions facilitated the resolution of conflicts, deliberation on criteria, and periodic adjustments to the review instructions. Updates and improvements were communicated through presentations and follow-up emails to ensure all team members remained informed. This iterative process of collaboration and feedback created a dynamic, transparent, and effective review workflow, ensuring alignment and shared understanding throughout the study.

2.4. Review management software

To manage the workflow and quality of the review process we decided at the stage of the project proposal to use a review management system. We have initially considered Rayyan as a possible tool. Some of the consortium members used this free on-line service [Ouzzani16] in former projects for systematic reviews [Fąferek24]. However, after a thorough analysis of its capabilities and the specific needs of the partner institutions, the WP2 coordinator recommended using a more advanced tool, Covidence.

The screening process for this rapid review was conducted using Covidence, a systematic review management platform selected for its robust features that enhance both efficiency and collaboration. Covidence streamlined every stage of the review process, from title and abstract screening to full-text assessment (Figure 1 serves as an example of the process of the full-text review). This enabled multiple

reviewers to work simultaneously, effectively resolving conflicts and improving the overall workflow. Key automation features, such as deduplication and PRISMA diagram generation, further accelerated the process and ensured transparency throughout. Its seamless integration with reference management tools also enhanced its utility, making it the ideal choice for this rapid review.

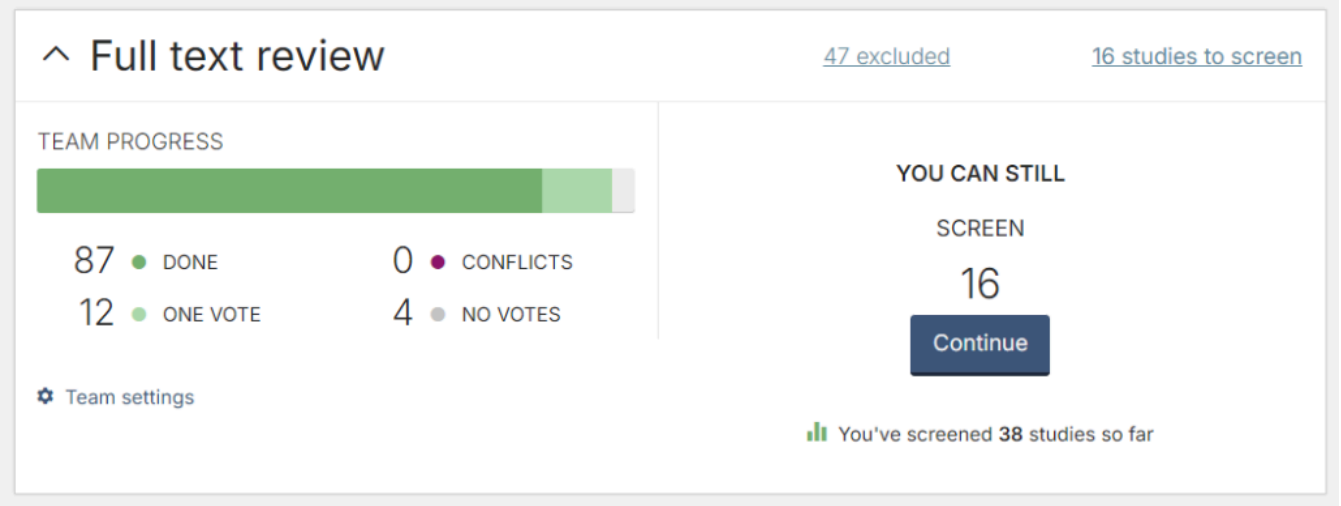


Figure 1: The full-text review process illustrated in Covidence

The selection of Covidence positively contributed to the quality of the review due to its anti-bias features, which help minimize errors and inconsistencies in data extraction and selection. The platform's systematic tracking of reviewer decisions and inclusion/exclusion criteria ensured a more reliable and reproducible review process. Additionally, as Covidence is part of the EMC research infrastructure, its use did not incur additional costs to the project.

2.5. Eligibility criteria

To guide the study selection process and facilitate the reliability of inclusions and exclusions, we established concise decision criteria. We conducted a pilot on the eligibility criteria using a sample of 30 publications. Based on the experiences with this pilot, we refined the eligibility criteria during the first alignment meeting with the rapid review team.

The description of the eligibility criteria was structured using the PICOS framework [Methley14]. The consecutive groups of inclusion and exclusion rules focused on:

- The Population (P) that were undergraduate medical and nursing students, and excluded were other professions and post-graduate students.
- The Intervention (I) which focused on the use of the five of the previously described categories D-CREDO tools, excluding studies from other technologies (e.g. virtual reality or 3D printing) or those where the use of the digital tool was just a medium to deliver standard content of clinical reasoning teaching.
- As according to the research question the review did not investigate alternative interventions, we skipped the Comparison (C) aspect.

- The Outcome (O) which was the educational content regarded as helpful in designing the learning units. These included learning objectives, educational or cognitive theories, learning activities, policies and guidelines, faculty development methods, and assessment methods related to the use of digital health tools in clinical reasoning.
- The Study design (S) for which we included original qualitative and quantitative papers, and perspective paper but excluded e.g. short conference abstracts, editorials, letters to editors and content not in English.

The detailed inclusion/exclusion criteria are presented in Appendix 2.

The eligibility criteria were later on implemented in the Covidence tool as a table that was easily referenced to during the review process (Figure 2).

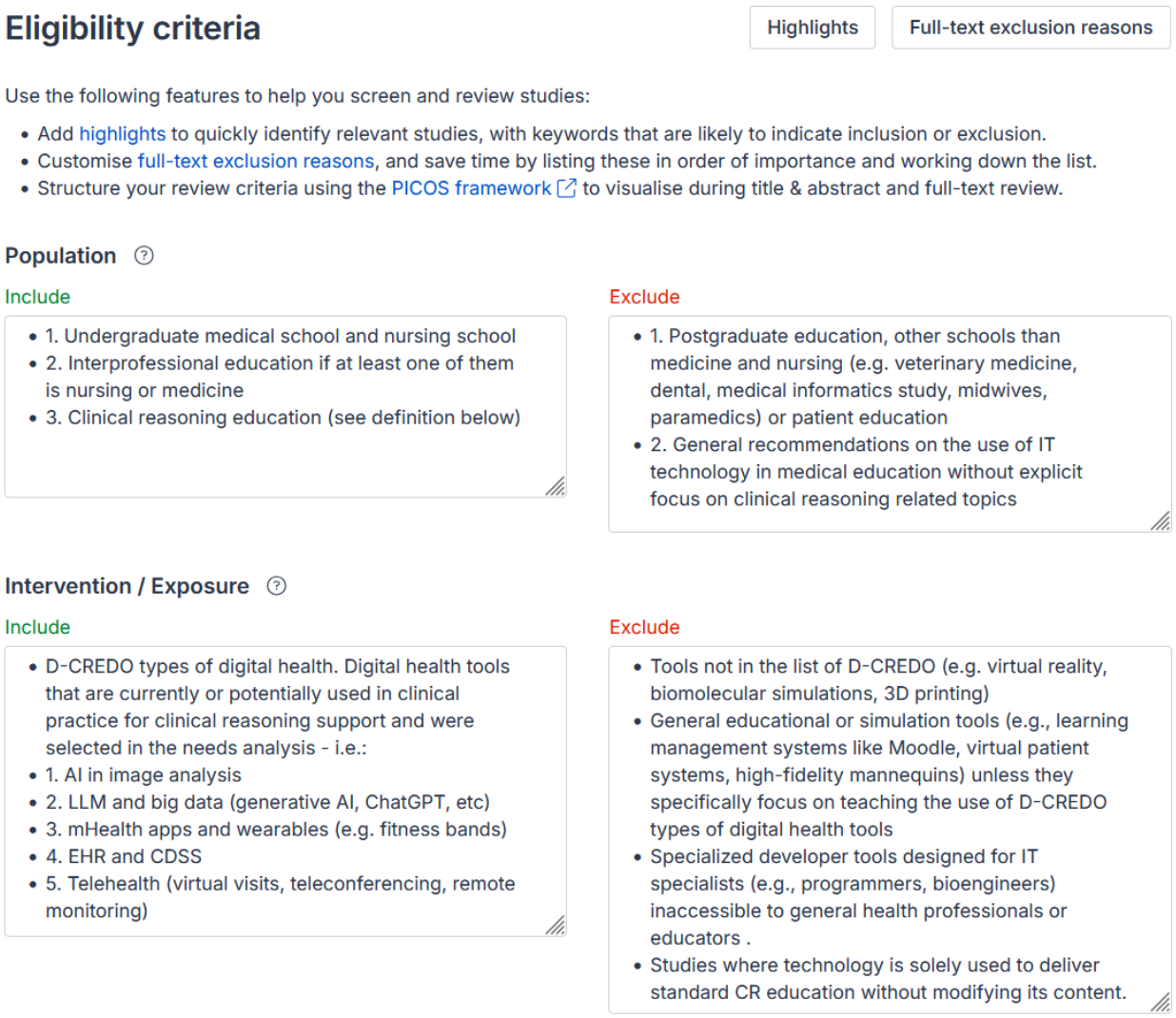


Figure 2: Eligibility criteria implemented as decision table in Covidence (part of the picture)

2.6. Data extraction form

The goal in the data extraction stage was to gather and organize data relevant to the integration of digital health tools into clinical reasoning education, which will serve as a foundation for the development of the D-CREDO curriculum. In particular we aimed to classify the papers into categories that would at the later stage facilitate easy information retrieval to build the learning objective catalogues, theories, viable learning activities, guidelines and policies. The process was designed to capture essential details about each study, including the study identification, study population, clinical reasoning focus, intervention/exposure details, educational use, and key findings.

Below is a breakdown of the data extraction categories and the corresponding information gathered from the studies:

- **Study Identification**: Basic information such as the title, authors, year of publication, country, and study design was recorded for each study, providing a clear overview of the research landscape.
- **Study Population**: Details about the total number of participants, the distinction between medical and nursing students, and the year in the educational program were captured. Additionally, the medical area or topic (e.g., surgery, general medicine) was noted to contextualize the studies.
- **Clinical Reasoning Focus**: The key stages of clinical reasoning (e.g., information gathering, hypothesis generation, differential diagnosis, management/treatment) were listed and identified based on their inclusion in each study.
- **Intervention/Exposure (D-CREDO Tools)**: Studies were categorized by the type of digital health tool used, such as AI in image analysis, LLM, mHealth apps, EHR, CDSS, and telehealth. Specific tools/products employed in each study were also documented.
- **Educational Use:** The educational components, including learning objectives, cognitive theories, learning activities, and faculty methods, were recorded to understand how digital tools were integrated into teaching and learning.
- **Key Findings**: The key findings summarized the results of each study, focusing on the effectiveness, challenges, and innovations of digital health tools in clinical reasoning education.

The detailed data extraction form used in this review is provided in Appendix 3.

3. Results

3.1. Abstract screening

To identify relevant studies for the rapid review, we conducted a comprehensive search across the five databases recommended by the qualified librarian. We performed the literature query on September 30, 2024. Included studies published after January 1, 2019 to September 30, 2024 (i.e. coverage of 5 years and 10 months).

The table 1 provides an overview of the database, platforms, and the number of records retrieved and included after duplicate removal from individual databases. No additional database limits were applied beyond those specified in the search strategies.

Table 1: Overview on social media platforms used by partners.

Database searched	Platform	Records	Records after duplicates removed
Medline ALL	Ovid	638	635
Embase	Embase.com	986	412
Web of Science Core Collection*	Web of Knowledge	485	291
PsycINFO	Ovid	69	42
ERIC	Ovid	60	47
Total		**2876**	**1527**

*Science Citation Index Expanded (1975-present) ; Social Sciences Citation Index (1975-present) ; Arts & Humanities Citation Index (1975-present) ; Conference Proceedings Citation Index- Science (1990-present) ; Conference Proceedings Citation Index- Social Science & Humanities (1990-present) ; Emerging Sources Citation Index (2005-present)

No other database limits were used than those specified in the search strategies

In total, 2,876 records were initially retrieved. The literature references were collected in RIS format and uploaded to EndNote for removal of duplicates. After the removal process, 1,527 unique entries were identified, including one record that was added manually.

The PRISMA flow of the identified abstracts after upload to Covidence presents Figure 2.

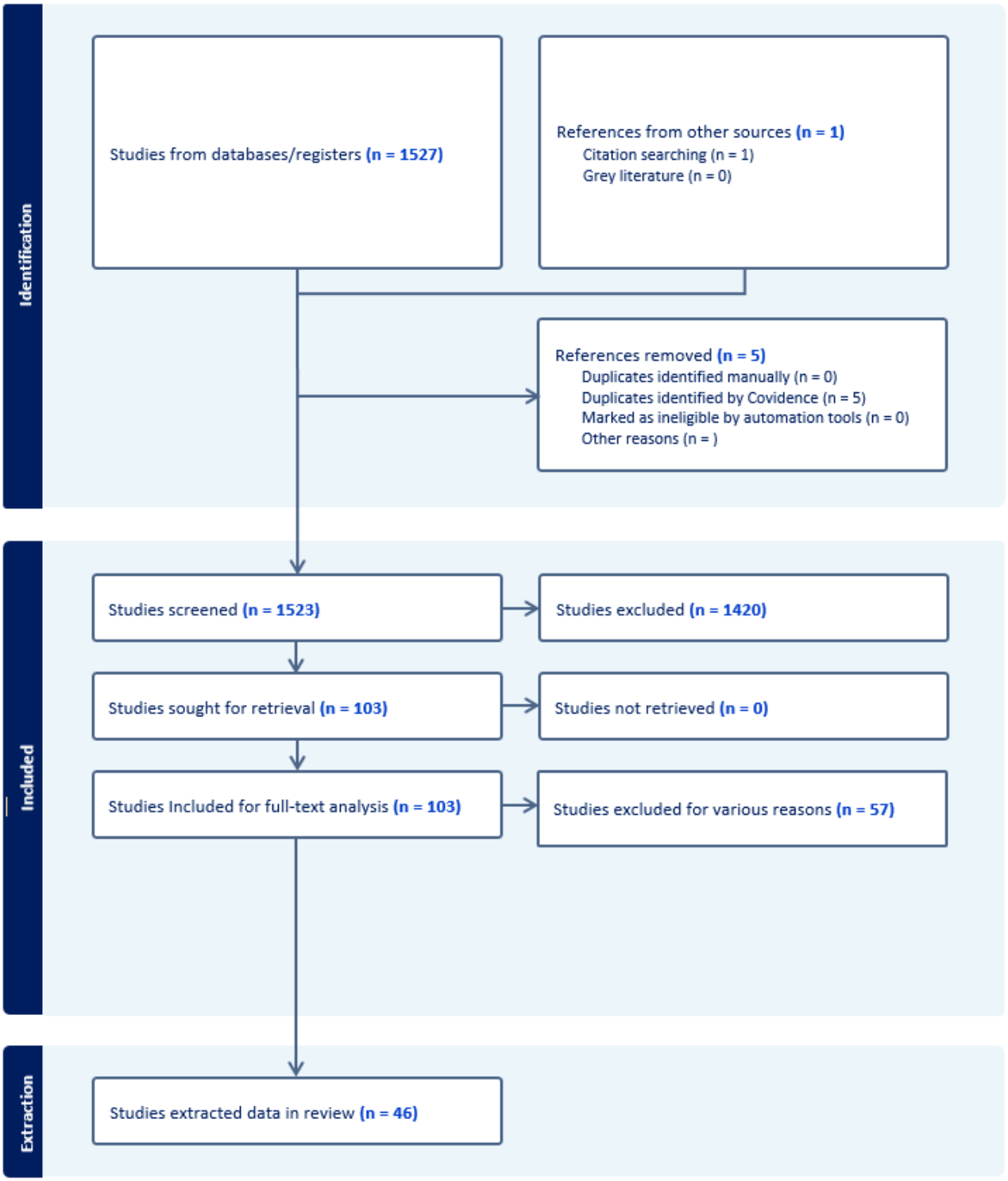


Figure 2: PRISMA flowchart of the study

3.2. Full-text analysis

The abstract screening process resulted in 103 studies included for full-text analysis and description. Thanks to a wide access to different full-text repositories across the project consortium and associate partners we were able to retrieve PDF files of all the included studies. These underwent a comprehensive full-text review to confirm their eligibility based on the predefined criteria. This detailed analysis involved a deeper examination of each study's alignment with the research objectives.

As a result, based on independent decision of two reviewers, mediated when needed by a third, of the 103 studies 57 studies were excluded from further stages of the review for various reasons:

- 24 studies involved the wrong population, such as postgraduate students or disciplines outside of medicine and nursing, which did not meet the focus on undergraduate medical and nursing students.
- 16 studies reported outcomes that were unrelated to the research question, either by addressing different topics or lacking a clear connection to clinical reasoning education and digital health tools.
- 7 studies investigated interventions or exposures outside the D-CREDO framework, which includes AI for image analysis, LLM, mHealth apps, EHR, CDSS, and telehealth, making them unsuitable for inclusion.
- 6 studies had unsuitable study characteristics, such as being in a non-English language or consisting of short conference abstracts that lacked sufficient detail for full analysis.
- 4 studies were excluded for other reasons, such as methodological issues or insufficient relevance to the research focus.

3.3. Data extraction results

The data extraction process for this review included 46 studies that met the predefined inclusion criteria. Over time, there has been a growing emphasis on Digital Health Technologies-Augmented Clinical Reasoning Education, with a notable increase in studies published each year. In 2019, only 1 study was included, but this number grew to 5 studies in 2020 and 2021. In 2022, 6 studies were published, and 2023 saw a jump to 11 studies. The most significant rise occurred in 2024, with 15 studies, in addition to 2 preprints. This clear upward trend highlights the expanding interest and research in this area.

The majority of studies focus on medical students, with a total of 29 studies. A smaller portion of the studies (11) focus on nursing education, and only 1 study explores both medical and nursing students. Additionally, 9 studies consider other populations.

Geographically, the studies are widely distributed, with North America contributing the largest number of studies (28), mainly from the USA (23 studies). Asia follows with 14 studies, and Europe contributes 13 studies. A smaller number of studies come from Oceania (2 studies) and South America (1 study from Brazil). Most of the research is conducted in high-income countries (HIC), including the USA, Canada, Germany, Switzerland, and Japan. While there is some representation from middle-income countries, such as Brazil and China, the majority of studies come from High-Income-Countries (HIC), particularly the USA.

In terms of study designs, the data reveals a broad range of methodologies used. The most common design is non-randomized experimental studies, which make up 9 of the studies. Randomized controlled trials (RCTs) are the second most frequent design (6 studies), followed by perspective papers (8 studies). Other types of research include qualitative studies (5), cohort studies (3), case-control studies (2), and a variety of reviews (systematic, scoping, and narrative). This diversity in study designs reflects the complexity and varied approaches taken to examine the role of digital health tools in clinical reasoning education.

When it comes to the types of digital health tools investigated, Large Language Models (LLMs) and big data are the most widely studied, appearing in 19 studies. These tools, such as AI-powered platforms like ChatGPT, are explored for their potential to enhance clinical reasoning. Electronic Health Records (EHR) follow closely behind, with 10 studies focusing on their integration into clinical reasoning education. Clinical Decision Support Systems (CDSS), AI in image analysis, and telehealth tools are less frequently studied, with 6, 4, and 3 studies respectively. A diverse range of other digital health tools, such as mobile health apps and specialized platforms, are covered in 7 studies.

The educational focus in these studies is also diverse, with learning activities being the most commonly addressed category, appearing in 28 studies. These learning activities include interactive simulations, case-based learning, and problem-solving exercises. The assessment of clinical reasoning (CR) is another important area, with 13 studies examining how digital tools are used to evaluate decision-making and reasoning. Faculty methods, such as blended learning or flipped classrooms, are addressed in 9 studies, and policies and guidelines related to the integration of digital tools into curricula and practice are discussed in 7 studies. Fewer studies focus on learning objectives (6) and educational or cognitive theories (4), suggesting that while the practical application of digital tools in clinical reasoning education is well-explored, theoretical and policy-oriented discussions remain less prevalent.

To summarize, the data extracted from these 46 studies shows a growing and diverse body of research focused on the integration of digital health technologies in clinical reasoning education. While there is strong emphasis on LLMs, EHRs, and learning activities, there are notably fewer articles about research related to telehealth, mobile apps, and AI in image analysis.

3. Discussion

Work Package 2 aims to clarify key components for the project's success, including defining terms, identifying educational theories and methods, proposing research questions for pilot studies, and guiding the integration of Learning Units (LUs) into curricula. Its goal is to establish a strong theoretical and methodological foundation for the project’s progress. This requires a longitudinal, systematic process of collecting sources and building an evidence database.

The first completed task under WP2 was the development of this report based on the rapid literature review. We pursued and achieved the key performance indicators defined at the stage of writing the D-CREDO proposal. In several cases, we exceeded the initial plans, contributing to a higher quality deliverable.

In particular, we ensured the design of a high-quality search strategy, incorporating several relevant keywords for each of the five D-CREDO types of digital tools. To further assure quality, we involved a trained librarian in designing the query. The search strategy was implemented in five research databases, which is

an unusually high number for a rapid review. The query covered almost six years of research (exceeding the required five years).

We had an excellent turnout of reviewers from the D-CREDO consortium for this task. While initially planning to involve five reviewers, we ended up with a team of 13 reviewers with diverse backgrounds, including medicine, nursing, educational science, and medical informatics. The work effort was equally distributed among the participating reviewers. Thanks to access to shared research infrastructure, we were able to use, at no additional cost, the professional review management tool Covidence, which includes several built-in features to reduce bias in the review process. The tool also allowed us to follow through a set of dashboards how the individual steps of the review progress.

We ensured that the search strategy, after deduplication, yielded over 1,000 abstracts as planned, ultimately screening more than 1,500 records. In the end we had 103 papers qualified for full-text analysis. Each of these was retrieved and analysed in its full-lengths parallel by two reviewers and either included in subsequent stages of data extraction or described with reasons for exclusion. A total of 46 studies were included for data extraction, while 56 were excluded with justifications.

In the included studies we were able to find good examples of papers describing educational theories, learning and assessment methods, digital tools in the context of CR practice and education for early and late stages of undergraduate curriculum and for both medicine and nurses health professions. The commented set of selected relevant papers is available as a spreadsheet for the later use by the consortium. Some of the characteristics of the studies followed the desired distribution - e.g. a balanced inclusion of studies for medical and nursing students. However, when it came to the distribution of types of D-CREDO tools we were surprised by the large number of reports on the use of ChatGPT and related large language models while other digital health tools like mobile apps received only little attention in the recent literature. This disparity is likely because mobile apps have been in use for over a decade, making them well-embedded in practice and less of a current research focus. In contrast, ChatGPT and large language models represent novel technologies that are still being extensively studied due to their recent emergence and perceived potential. To ensure the development of high-quality learning units for all selected D-CREDO categories of tools we need to broaden our search as part of the development of learning units for content in other sources like grey literature, contact with experts in the consortium or associated partners to collect the necessary sources. Additionally, it might be worth extending our search timeframe to include older, potentially significant articles on mobile apps from the earlier stages of their adoption, as these may provide valuable insights.

Moving forward, the data gathered from this review will inform the creation of the learning objectives and play a crucial role in shaping the project's conceptual framework which are the immediate following activities from this deliverable.

4. Summary and conclusions

The first task completed under WP2 involved creating a literature repository to guide the educational framework. This activity was conducted through a rapid review involving many consortium members with diverse backgrounds. The review identified valuable studies, which were indexed for retrieving useful

information about educational components, including learning objectives, cognitive theories, learning activities, and faculty methods. However, it also highlighted gaps in the representation of certain D-CREDO tools in clinical reasoning education within the review results. The insights gained from the rapid review will play a key role in shaping the development of the D-CREDO curriculum.

5. References

[Faferek24] Fąferek J, Cariou PL, Hege I, Mayer A, Morin L, Rodriguez-Molina D, Sousa-Pinto B, Kononowicz AA. Integrating virtual patients into undergraduate health professions curricula: a framework synthesis of stakeholders' opinions based on a systematic literature review. BMC Med Educ. 2024 Jul 5;24(1):727.

[Gough17] Gough D, Thomas J, Oliver S, editors. An introduction to systematic reviews. 2nd ed. London: SAGE Publications; 2017.

[Huesmann23] Huesmann L, Sudacka M, Durning SJ, Georg C, Huwendiek S, Kononowicz AA, Schlegel C, Hege I. Clinical reasoning: What do nurses, physicians, and students reason about. J Interprof Care. 2023 Nov 2;37(6):990-998.

[Methley14] Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res. 2014 Nov 21;14:579.

[Ouzzani16] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016 Dec 5;5(1):210.

[Tricco17] Tricco AC, Langlois E, Straus SE, World Health Organization. Rapid reviews to strengthen health policy and systems: a practical guide. World Health Organization; 2017.

[Young19] Young M, Thomas A, Gordon D, Gruppen L, Lubarsky S, Rencic J, Ballard T, Holmboe E, Da Silva A, Ratcliffe T, Schuwirth L, Durning SJ. The terminology of clinical reasoning in health professions education: Implications and considerations. Med Teach. 2019 Nov;41(11):1277-1284.

6. Appendices

Appendix 1 - Database search – Teaching digital technologies medical education

This appendix details the comprehensive database search conducted to identify studies related to teaching digital health technologies in undergraduate medical and nursing education.

Medline

(exp * Artificial Intelligence/ OR * Radiomics/ OR * Natural Language Processing/ OR * Big Data/ OR * Data Mining/ OR * exp Mobile Applications/ OR * exp Machine Learning/ OR * Smartphone/ OR * Wearable Electronic Devices/ OR *

Biosensing Techniques/ OR * Electronic Health Records/ OR * Telemedicine/ OR * Remote Sensing Technology/ OR * Automation/ OR * Decision Support Systems, Clinical / OR ((artificial* ADJ3 intelligen*) OR generative-ai OR ((convolution* OR artificial*) ADJ3 neural-network*) OR radiomic* OR computer-vision* OR natural-language-process* OR prompt-engineer* OR ((machine OR deep) ADJ3 Learning) OR (large-languag* ADJ3 model*) OR chatgpt OR chat-gpt OR gpt4 OR gpt-4 OR gpt3* OR gpt-3* OR transformer*-architecture* OR openai OR open-ai OR ((claude OR anthropic OR gemini) AND ai) OR text-generation* OR shot-learning OR big-data OR data-mining OR predictive-analytic* OR mhealth OR mobile-health OR ((mobile OR health) ADJ3 (app OR apps OR application*)) OR smartphone* OR smart-phone* OR iphone* OR i-phone* OR (digital* ADJ3 (sensor* OR monitor* OR phenotyp* OR marker*)) OR fitness-tracker* OR smartwatch* OR apple-watch* OR smart-textil* OR smart-ring* OR wearable* OR biosensor* OR bio*-sensor* OR ((electronic* OR digital*) ADJ3 (health OR patient* OR medical*) ADJ3 (record*)) OR symptom-checker* OR diagnosis-generator* OR telehealth* OR tele-health* OR ehealth* OR e-health* OR telecare* OR tele-care* OR telehealthcare* OR tele-healthcare* OR tele-health-care* OR telemedicine* OR tele-medicine* OR (virtual* ADJ3 (visit* OR diagnos*)) OR (remote ADJ3 (sensing OR monitoring)) OR (video ADJ3 consult*) OR (computer* ADJ3 (decision-support*)) OR ehr OR emr OR cdss OR (health ADJ3 monitoring)).ab,ti,kw. OR (ai OR app OR apps OR ehr OR emr OR automation* OR automated* OR neural-network*).ti.) AND (Education, Medical / OR Students, Medical / OR Education, Nursing / OR Schools, Medical / OR * Teaching / OR (((medical* OR nurse* OR nursing*) ADJ3 (educat* OR student* OR teaching OR graduate* OR university* OR school* OR curriculum*))).ab,ti,kw. OR (education* OR student* OR teaching* OR curriculum*).ti.) AND (Clinical Reasoning/ OR Clinical Decision-Making/ OR clinical decision making/ OR Decision Support Systems, Clinical / OR (((clinic* OR diagnos* OR therap* OR medical* OR management* OR critical*) ADJ3 (reasoning* OR decision* OR thinking* OR judgment* OR problem-solv*))).ab,ti,kw.) AND 2019:2030.(sa_year).

Embase

('artificial intelligence'/mj/exp OR 'computer vision'/mj OR radiomics/mj/exp OR 'natural language processing'/mj OR 'prompt engineering'/mj OR 'big data'/mj OR 'data mining'/mj/exp OR 'predictive analytics'/mj OR mhealth/mj OR 'mobile application'/mj OR 'machine learning'/mj/exp OR 'large language model'/mj/exp OR smartphone/mj OR 'wearable device'/mj OR 'digital phenotyping'/mj OR biosensor/mj OR 'electronic health record'/mj/exp OR 'electronic medical record'/mj OR telehealth/mj OR 'virtual visit'/mj OR telecare/mj OR telemedicine/mj OR 'remote sensing'/mj OR 'video consultation'/mj OR automation/mj OR 'clinical decision support system'/mj OR ((artificial* NEAR/3 intelligen*) OR generative-ai OR ((convolution* OR artificial*) NEAR/3 neural-network*) OR radiomic* OR computer-vision* OR natural-language-process* OR prompt-engineer* OR ((machine OR deep) NEAR/3 Learning) OR (large-languag* NEAR/3 model*) OR chatgpt OR chat-gpt OR gpt4 OR gpt-4 OR gpt3* OR gpt-3* OR transformer*-architecture* OR openai OR open-ai OR ((claude OR anthropic OR gemini) AND ai) OR text-generation* OR shot-learning OR big-data OR data-mining OR predictive-analytic* OR mhealth OR mobile-health OR ((mobile OR health) NEAR/3 (app OR apps OR application*)) OR smartphone* OR smart-phone* OR iphone* OR i-phone* OR (digital* NEAR/3 (sensor* OR monitor* OR phenotyp* OR marker*)) OR fitness-tracker* OR smartwatch* OR apple-watch* OR smart-textil* OR smart-ring* OR wearable* OR biosensor* OR bio*-sensor* OR ((electronic* OR digital*) NEAR/3 (health OR patient* OR medical*) NEAR/3 (record*)) OR symptom-checker* OR diagnosis-generator* OR telehealth* OR tele-health* OR ehealth* OR e-health* OR telecare* OR tele-care* OR telehealthcare* OR tele-healthcare* OR tele-health-care* OR telemedicine* OR tele-medicine* OR (virtual* NEAR/3 (visit* OR diagnos*)) OR (remote NEAR/3 (sensing OR monitoring)) OR (video NEAR/3 consult*) OR (computer* NEAR/3 (decision-support*)) OR ehr OR emr OR cdss OR (health NEAR/3 monitoring)):ab,ti,kw OR (ai OR app OR apps OR ehr OR emr OR automation* OR automated* OR neural-network*):ti) AND ('medical education'/exp OR 'medical student'/exp OR 'nursing education'/exp OR 'medical school'/de OR teaching/mj OR (((medical* OR nurse* OR nursing*) NEAR/3 (educat* OR student* OR teaching OR graduate* OR university* OR school* OR curriculum*))):Ab,ti,kw OR (education* OR student* OR teaching* OR curriculum*):ti) AND ('clinical reasoning'/exp OR 'diagnostic reasoning'/exp OR 'clinical decision making'/exp OR 'clinical decision support system'/de OR (((clinic* OR diagnos* OR therap* OR medical* OR management* OR critical*) NEAR/3 (reasoning* OR decision* OR thinking* OR judgment* OR problem-solv*))):Ab,ti,kw) AND [2019-2024]/py NOT [conference abstract]/lim

Web of science

(TS=(((((artificial* NEAR/2 intelligen*) OR generative-ai OR ((convolution* OR artificial*) NEAR/2 neural-network*) OR radiomic* OR computer-vision* OR natural-language-process* OR prompt-engineer* OR ((machine OR deep) NEAR/2 Learning) OR (large-languag* NEAR/2 model*) OR chatgpt OR chat-gpt OR gpt4 OR gpt-4 OR gpt3* OR gpt-3* OR transformer*-architecture* OR openai OR open-ai OR ((claude OR anthropic OR gemini) AND ai) OR text-generation* OR shot-learning OR big-data OR data-mining OR predictive-analytic* OR mhealth OR mobile-health OR ((mobile OR health) NEAR/2 (app OR apps OR application*)) OR smartphone* OR smart-phone* OR iphone* OR i-phone* OR (digital* NEAR/2 (sensor* OR monitor* OR phenotyp* OR marker*)) OR fitness-tracker* OR smartwatch* OR apple-watch* OR smart-textil* OR smart-ring* OR wearable* OR biosensor* OR bio*-sensor* OR ((electronic* OR digital*) NEAR/2 (health OR patient* OR medical*) NEAR/2 (record*)) OR symptom-checker* OR diagnosis-generator* OR telehealth* OR tele-health* OR ehealth* OR e-health* OR telecare* OR tele-care* OR telehealthcare* OR tele-healthcare* OR tele-health-care* OR telemedicine* OR tele-medicine* OR (virtual* NEAR/2 (visit* OR diagnos*)) OR (remote NEAR/2 (sensing OR monitoring)) OR (video NEAR/2 consult*) OR (computer* NEAR/2 (decision-support*)) OR ehr OR emr OR cdss OR (health NEAR/2 monitoring))) OR TI=(ai OR app OR apps OR ehr OR emr OR automation* OR automated* OR neural-network*))) AND (TS=(((medical* OR nurse* OR nursing*) NEAR/2 (educat* OR student* OR teaching OR graduate* OR university* OR school* OR curriculum*))) OR TI=(education* OR student* OR teaching* OR curriculum*)) AND TS=((((clinic* OR diagnos* OR therap* OR medical*) NEAR/2 (reasoning* OR decision* OR thinking* OR judgment* OR problem-solv*)) OR cdss OR patient-portal*) OR (decision-mak*)) AND PY=(2019-2030) AND LA=(english) AND DT=(article)

PsycINFO

(exp * Artificial Intelligence/ OR * Natural Language Processing/ OR * Big Data/ OR * Data Mining/ OR * exp Mobile Applications/ OR * exp Machine Learning/ OR * Smartphones/ OR * Wearable Devices/ OR * Electronic Health Records/ OR * Telemedicine/ OR * Automation/ OR ((artificial* ADJ3 intelligen*) OR generative-ai OR ((convolution* OR artificial*) ADJ3 neural-network*) OR radiomic* OR computer-vision* OR natural-language-process* OR prompt-engineer* OR ((machine OR deep) ADJ3 Learning) OR (large-languag* ADJ3 model*) OR chatgpt OR chat-gpt OR gpt4 OR gpt-4 OR gpt3* OR gpt-3* OR transformer*-architecture* OR openai OR open-ai OR ((claude OR anthropic OR gemini) AND ai) OR text-generation* OR shot-learning OR big-data OR data-mining OR predictive-analytic* OR mhealth OR mobile-health OR ((mobile OR health) ADJ3 (app OR apps OR application*)) OR smartphone* OR smart-phone* OR iphone* OR i-phone* OR (digital* ADJ3 (sensor* OR monitor* OR phenotyp* OR marker*)) OR fitness-tracker* OR smartwatch* OR apple-watch* OR smart-textil* OR smart-ring* OR wearable* OR biosensor* OR bio*-sensor* OR ((electronic* OR digital*) ADJ3 (health OR patient* OR medical*) ADJ3 (record*)) OR symptom-checker* OR diagnosis-generator* OR telehealth* OR tele-health* OR ehealth* OR e-health* OR telecare* OR tele-care* OR telehealthcare* OR tele-healthcare* OR tele-health-care* OR telemedicine* OR tele-medicine* OR (virtual* ADJ3 (visit* OR diagnos*)) OR (remote ADJ3 (sensing OR monitoring)) OR (video ADJ3 consult*) OR (computer* ADJ3 (decision-support*)) OR ehr OR emr OR cdss OR (health ADJ3 monitoring)).ab,ti. OR (ai OR app OR apps OR ehr OR emr OR automation* OR automated* OR neural-network*).ti.) AND (Medical Education/ OR Medical Students/ OR Nursing Education/ OR Medical Schools/ OR * Teaching / OR (((medical* OR nurse* OR nursing*) ADJ3 (educat* OR student* OR teaching OR graduate* OR university* OR school* OR curriculum*))).ab,ti. OR (education* OR student* OR teaching* OR curriculum*).ti.) AND ((((clinic* OR diagnos* OR therap* OR medical*) ADJ3 (reasoning* OR decision* OR thinking* OR judgment* OR problem-solv*)) OR cdss OR patient-portal*).ab,ti. OR (decision-mak*).ti.) AND 2019:2030.(sa_year).

ERIC

(exp Artificial Intelligence/ OR Natural Language Processing/ OR * Automation/ OR ((artificial* ADJ3 intelligen*) OR generative-ai OR ((convolution* OR artificial*) ADJ3 neural-network*) OR radiomic* OR computer-vision* OR natural-language-process* OR prompt-engineer* OR ((machine OR deep) ADJ3 Learning) OR (large-languag* ADJ3 model*) OR chatgpt OR chat-gpt OR gpt4 OR gpt-4 OR gpt3* OR gpt-3* OR transformer*-architecture* OR openai OR open-ai OR ((claude OR anthropic OR gemini) AND ai) OR text-generation* OR shot-learning OR big-data OR

data-mining OR predictive-analytic* OR mhealth OR mobile-health OR ((mobile OR health) ADJ3 (app OR apps OR application*)) OR smartphone* OR smart-phone* OR iphone* OR i-phone* OR (digital* ADJ3 (sensor* OR monitor* OR phenotyp* OR marker*)) OR fitness-tracker* OR smartwatch* OR apple-watch* OR smart-textil* OR smart-ring* OR wearable* OR biosensor* OR bio*-sensor* OR ((electronic* OR digital*) ADJ3 (health OR patient* OR medical*) ADJ3 (record*)) OR symptom-checker* OR diagnosis-generator* OR telehealth* OR tele-health* OR ehealth* OR e-health* OR telecare* OR tele-care* OR telehealthcare* OR tele-healthcare* OR tele-health-care* OR telemedicine* OR tele-medicine* OR (virtual* ADJ3 (visit* OR diagnos*)) OR (remote ADJ3 (sensing OR monitoring)) OR (video ADJ3 consult*) OR (computer* ADJ3 (decision-support*)) OR ehr OR emr OR cdss OR (health ADJ3 monitoring)).ab,ti. OR (ai OR app OR apps OR ehr OR emr OR automation* OR automated* OR neural-network*).ti.) AND ((((clinic* OR diagnos* OR therap* OR medical*) ADJ3 (reasoning* OR decision* OR thinking* OR judgment* OR problem-solv*)) OR cdss OR patient-portal*).ab,ti. OR (decision-mak*).ti.) AND 2019:2030.(sa_year).

Appendix 2 - Inclusion/Exclusion Criteria

Population

Include

- Undergraduate medical school and nursing school
- Interprofessional education if at least one of them is nursing or medicine
- Clinical reasoning education (see definition below)

Exclude

- Postgraduate education, other schools than medicine and nursing (e.g. veterinary medicine, dental, medical informatics study, midwives, paramedics) or patient education
- General recommendations on the use of IT technology in medical education without explicit focus on clinical reasoning related topics

Intervention / Exposure

Include

- D-CREDO types of digital health. Digital health tools that are currently or potentially used in clinical practice for clinical reasoning support and were selected in the needs analysis - i.e.:
- AI in image analysis
- LLM and big data (generative AI, ChatGPT, etc)
- mHealth apps and wearables (e.g. fitness bands)
- EHR and CDSS
- Telehealth (virtual visits, teleconferencing, remote monitoring)

Exclude

- Tools not in the list of D-CREDO (e.g. virtual reality, biomolecular simulations, 3D printing)
- General educational or simulation tools (e.g., learning management systems like Moodle, virtual patient systems, high-fidelity mannequins) unless they specifically focus on teaching the use of D-CREDO types of digital health tools
- Specialized developer tools designed for IT specialists (e.g., programmers, bioengineers) inaccessible to general health professionals or educators .

Outcome

Include, reports on:

- Learning objectives for the use of D-CREDO types of digital health tools in clinical reasoning
- Educational or cognitive theories that explain or make recommendation on how to use D-CREDO types of digital health tools in clinical reasoning
- Learning activities around the use of D-CREDO types of digital health tools in clinical reasoning
- Policies and guidelines for meaningful, safe, ethical use of D-CREDO types in clinical reasoning education and teaching of use in practice

- Methods on how to use D-CREDO tools by faculty at medical or nursing schools to design or conduct clinical reasoning teaching activities (e.g. prompt engineering methods in deliberate practice of clinical reasoning)

Exclude
- Effectiveness studies benchmarking digital health tools against humans (e.g. ability of ChatGPT to answer exam questions tested against students)
- Technical reports of how D-CREDO types of digital health tools or educational were developed or how they work from the technical perspective (e.g. what algorithms were implemented)
- Effectiveness (does it work) studies of proprietary digital health and educational tools in clinical reasoning, unless the goal is to design and evaluate a generalizable educational activity around that type of tool.
- Readiness studies reporting on willingness or (technical) preparedness to learn about digital health tools are excluded, unless the study addresses the need for specific learning objectives.

Study Characteristics
Include
- Original research both qualitative and quantitative
- Perspective papers
- Reviews of different types (narratives, scoping)

Exclude
- Short conference abstracts
- Editorials
- Letters to the editor
- Articles not in English

Others
Include
Definition Clinical reasoning education
"Clinical Reasoning encompasses health professionals thinking and acting in assessment, diagnostic, and management processes in clinical situations taking into account the patient's specific circumstances and preferences" DID-ACT definition: [Huesman23].

Clinical reasoning includes diagnostic reasoning (e.g., problem representation, selection and interpretation of diagnostic tests, differential diagnosis, and decision on diagnosis, diagnostic error prevention) and management reasoning (e.g., a patient-centered approach to treatment/care selection, including cultural and ethical aspects). Communication skills and physical examination skills are a borderline case; included are studies that deal with the strategic acquisition of patient information relevant to the medical problem. Interprofessional communication is included when focused on collaborative decision-making related to diagnosis or patient care.

Appendix 3 – Data extraction format

Study Identification

Title

Title of paper/abstract/report from which data are extracted: *(...)*

Author(s)

First author only: *(...)*

Year of Publication

· 2024

- 2023
- 2022
- 2021
- 2020
- 2019
- Preprint
- Other: *(...)*

Country

Country in which the study was conducted: *(...)*

Study Design

- Randomised controlled trial
- Non-randomised experimental study
- Cohort study
- Cross-sectional study
- Case-control study
- Systematic review
- Qualitative research
- Perspective papers
- Case series/reports
- Protocols
- Other: *(...)*

Notes

(...)

Study Population

Population

- Medical students
- Nursing students
- Both
- Other: *(...)*

Total Number of Participants

Number: *(...)*

Number of Medical and/or Nursing Students Separately

Medical: *(...)*

Nursing: *(...)*

Year in the Educational Program

Year: *(...)*

Medical Area/Topic

Specify the medical area or topic the study focuses on (e.g., surgery, internal medicine, general medicine):

(...)

Notes

(...)

Clinical Reasoning (From Daniel et al., 2019)

- Information gathering
- Hypothesis generation
- Problem representation
- Differential diagnosis
- Leading or working diagnosis
- Diagnostic justification
- Management and treatment
- General (not specified)
- Other: *(...)*

Notes

(...)

Intervention/Exposure (D-CREDO Tools)

Type of Digital Health Tool

- AI in image analysis
- Large Language Models (LLMs) & Big Data (e.g., generative AI, ChatGPT)
- mHealth Apps & Wearables (e.g., fitness trackers)
- Electronic Health Records (EHR)
- Clinical Decision Support Systems (CDSS)
- Telehealth (virtual visits, teleconferencing, remote monitoring)
- Other: *(...)*

What Specific Product/Tool Was Used?

(...)

Notes

(...)

Educational Use

Type of Educational Use

- Learning objectives
- Educational or cognitive theories
- Learning activities
- Policies and guidelines

- Faculty methods
- Assessment of clinical reasoning (CR)
- Other: *(...)*

Details

(...)

Notes

(...)

Key Findings

Results

Brief description of the results: *(...)*

Conclusions

Study's final conclusions and relevance to clinical reasoning education: *(...)*

Notes

(...)