European Journal of Inclusive Education (EJIE)
2026 Vol. 5, Issue 1,
https://doi.org/10.7146/ejie.v5i1.158034

AI-driven systematic reviews under scrutiny: Is AI-driven literature screening an alternative to human screening? A synopsis of three systematic reviews on the use of ICT in inclusive school settings

Claudia Mertens[1], Carolin Quenzer-Alfred, Anna-Maria Kamin, Tim Homrighausen, Daniel Mays

Abstract

In the age of digitization, societal participation is increasingly tied to media competence, making the acquisition of media competence a crucial educational objective - especially within inclusive and special educational needs school settings. As schools – among other stakeholders in education - are responsible for fostering media literacy, educators face the question of which didactic digital approaches help to support inclusion. Empirical research is needed to further leverage the potential benefits/risks of digital media in inclusive/SEN settings. However, empirical research must be preceded by a comprehensive understanding of existing studies on the intersection of digital media and inclusive/SEN education. This necessitates systematic literature reviews, which are traditionally conducted through manual screening. Given the time-intensive nature of this task, AI-assisted screening tools have gained attention as a possible alternative. Yet, AI-assisted screening tools require careful evaluation regarding accuracy and bias. By comparing manual and AI-driven screening processes, this study highlights the need for a balanced approach that integrates technological advancements with human expertise. Therefore, this paper pursues two primary objectives. First, it aggregates the findings of three systematic literature reviews examining the role of digital media in supporting students with special educational needs in primary and secondary education. These reviews included peer-reviewed and non-peer-reviewed empirical studies published in German and English over the past ten years. Criteria focused explicitly on digital media use in inclusive or special education contexts, excluding tertiary education. Second, the paper explores the methodological limitations of aggregating the results of three different reviews and it focusses on implications of using AI for literature screening by comparing the outcomes of manual versus AI-assisted approaches. While AI tools offer efficiency, serious concerns remain - regarding the accuracy of machine decisions and the potential introduction of new biases. The replication of one of the above-mentioned three reviews using AI allows for a direct comparison, highlighting the advantages and limitations of automated screening. This contribution aims to inform both the research community and educators by presenting current findings on digital media and inclusion while also critically reflecting on the evolving role of AI in academic research processes.

Keywords: Digital Media, Inclusive Education, Special Educational Needs, Systematic Review, AI-Screening

Points of interest

- The first (but subordinate) aim of this article is to sum up results of three systematic reviews on the interlink between inclusion/SEN (special educational needs) and digital media in school

[1] Corresponding author: claudia.mertens@uni-bielefeld.de

ISSN: 2794-4417
https://tidsskrift.dk/ejie

settings. Readers who are interested in more details are referred to the other publications of the author team (see references).

- The overall key finding of this screening process of the three reviews was that digital optimism seems to be justified. Some valuable research is available on the didactic use of digital media but more research is needed in the area of the "acquisition" of media literacy (in the sense of "learning *about* media"). More studies from a qualitative research paradigm are necessary.
- The second aim of this paper, which is the focus of this article, has a methodological perspective. The objective is to work out methodological challenges of the aggregation of the results of the three reviews and, first and foremost, to compare AI-driven literature screening to manual screening processes.
- The key finding of this second methodological research question is that AI driven screening is extremely helpful but cannot fully replace human screening.

Introduction

In the culture of digitization (Stalder, 2016) societal participation is bound to media competence (Bosse et al., 2019). In this publication the term *media competence* is used in a broad definition: According to Baacke (2001) the concept encompasses *Medienkritik* ("reflecting media critically"), *Medienkunde* ("knowledge about ICT"), *Mediennutzung* ("using ICT") and *Mediengestaltung* ("media design") [translation by the authors]. In a wide understanding of the definition by Baacke this definition already acknowledges the multiple interweaving of the subject with the world. In Germany, nevertheless, quite often scientists give preference to the term *Medienbildung* as the term claims to go beyond *Medienkompetenz* by giving respect to the "self's constant entanglement with the world" (Pieper et al. 2023: 62). As Pieper et al. (2023) elaborate, there is no direct equivalent to *Medienbildung* in the international discourse. The discussion of the terminology would go beyond the scope of this paper. Yet, the reader has to keep in mind that when we speak of *media competence* or *media education* we mean the holistic concept - referring to the ability to access, analyze, evaluate, and create media in various forms.

As participation is bound to media competence, pupils have to acquire media competence along the line of their school biography. School settings are responsible for providing media education. Likewise, teachers have to teach media competence for the sake of fostering inclusion. The question is delicate, which didactic setting is best to profit from mediatization for the aim of inclusion. Yet, before designing research projects on the challenges and benefits of digital media for inclusive school settings, it is – as a preliminary step – mandatory to grasp the state of the already existing research in the scientific community. It is only afterwards, that consequences for didactic settings can be drawn.

The aim of getting such a general survey of empirical research (on the interlink of digitization and inclusion in this case) can be reached by doing so-called literature reviews or scoping reviews. But screening the literature manually is an extremely time-consuming process and is, consequently, getting more and more automated with the help of AI-driven screening software. The question arises as to whether it would be advisable to outsource the challenging and time-consuming task of literature screening to machine analysis. Before using screening software, scientists need to discuss where possible flaws of AI can be identified - in comparison to the flaws caused by human bias. It is only after thorough assessment of the pros and cons of AI-driven screening that scientific communities can rely on AI-software for replacing human expertise.

Having made this point, the aim of this paper is twofold: 1) Firstly, it shall briefly aggregate the results on three systematic reviews on the interlink between "digital media" and "inclusion" in school settings. 2) Secondly, it shall present in detail the pros and cons of AI-driven screening in comparison to manual screening made during the process of the three reviews.

Outline of the paper:

(1) In order to understand the methodological discussion in the second part of the paper it is, as mentioned above, necessary to give a brief summary of the three reviews. The research question of the first section of this paper is: "What is the current state of German-language/English-language research on digital media for students with additional or special educational needs?". The inclusion criteria were: publication in the past 10 years; inclusion of peer reviewed studies only (except for the second German review where the focus was deliberately on non-peer reviewed studies); publication language was German (reviews I and II) and English (review III); only empirical studies; focus on digital media, focus on special education and/or inclusion; primary or secondary schools setting; exclusion of research from the tertiary context; inclusion of research from the perspective of the students. In this paper the aggregation of all three reviews shall be presented[2].

(2) The methodologically-oriented research questions of the second main section of this paper are: "What are the methodological challenges of aggregating results of three reviews and what are the benefits and risks of AI-supported screening?". The assumption is that AI-driven screening might be extremely helpful (on condition that the results of AI screening were reliable), since a single data-base search can yield thousands of articles that must be reviewed for relevance. In a classical systematic review, the initial step after merging the results from multiple databases, is to retrieve the abstracts and then to de-duplicate the corpus. When starting the manual screening process, the data comprises a huge number of unlabeled records. Yet, due to the skyrocketing number of publications in the digital age it is highly time consuming to screen manually. Therefore, screening with the help of data-driven approaches may be promising. Nevertheless, before applying AI screening on a large scale, the validity of automatization needs a critical check.

Aggregation of the results of the manual screening process

In order to better grasp the methodological discussion of the paper in section II a brief overview of the manual screening process shall be given. Let us start with the first research question on what we know about the use of digital media for inclusive/SEN settings. To answer this question, first of all, the search string shall be disclosed, then the number of hits shall be revealed. In the following section the results shall be laid open. As mentioned at the beginning, the aim of the international review (= review III) was to replicate two systematic reviews of studies published in German (see Mertens et al., 2022; Quenzer-Alfred et al., 2023). The research string was more or less transferred identically in all three reviews – with only minor culturally necessary adaptions on the linguistic level:

(DIGITAL* OR MEDIA* OR TECHNOLO* OR VIRTUA* OR "AUGMENTED REALIT*" OR COMPUTER* OR TABLET OR IPAD OR "APP" OR "LEARNING MANAGEMENT SYSTEM" OR INTERNET* OR SOFTWARE OR MOBILE OR CYBER OR GAME OR ICT OR "E-LEARNING" OR "DISTANCE EDUCATION" OR "WEB-BASED LEARNING" OR "ONLINE LEARNING" OR "WEB-BASED TRAINING" OR "ONLINE EDUCATION" OR "WEB-BASED EDUCATION") AND ("ELEMENTARY SCHOOL" OR "PRIMARY SCHOOL" OR "SECONDARY SCHOOL" OR "ELEMENTARY EDUCATION" OR "PRIMARY EDUCATION" OR "SECONDARY EDUCATION" OR PUPIL* OR STUDENT* OR LEARN* OR "K-12" OR "K12") AND ("SPECIAL NEED*" OR "SEN" OR "SPECIAL EDUCATION*" OR "SPECIAL SUPPORT NEED*" OR „ADDITIONAL LEARNING NEED*" OR "COMPLEX NEED*" OR "ADDITIONAL SUPPORT NEED*" OR "ADDITIONAL NEED*" OR ASN OR "SPECIAL SCHOOL" OR IMPAIR* OR HANDICAP* OR DISABLE* OR DISABILIT* OR DISORDER OR DIFFICULTY OR DIFFICULTIES OR DEFICIT* OR

[2] For a more detailed description of the methodology and the findings from the review with peer-reviewed studies go to Quenzer-Alfred et al. (2023) (=review I); for a detailed description of the aggregated results of the studies with and without peer review from the German speaking context go to Mertens et al. (2022) (=review II); for a detailed description of the studies with peer review published in English go to Mertens et al. (n.d.; under review) (=review III).

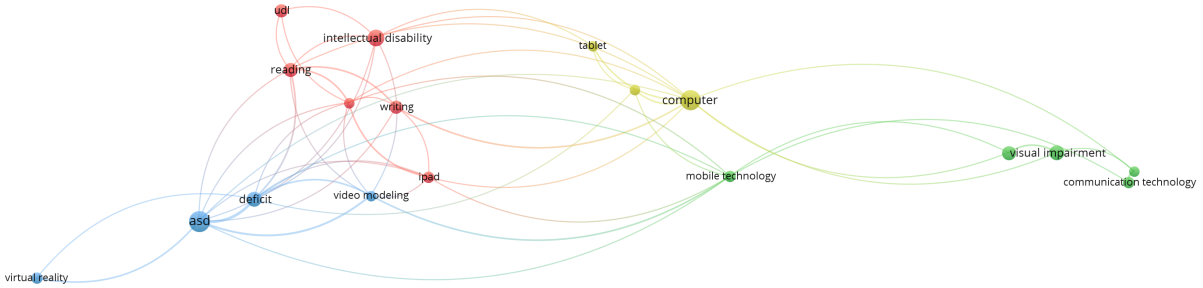
"EMOTIONAL DEVELOPMENT" OR "SOCIAL DEVELOPMENT" OR ADHD OR AUTIS* OR "INTELLECTUAL DEVELOPMENT" OR "SENSORY DEVELOPMENT" OR "MOTOR DEVELOPMENT" OR "PHYSICAL DEVELOPMENT" OR "LANGUAGE DEVELOPMENT" OR HEAR* OR DEAF* OR BLIND* OR INCLUSI* OR BARRIER OR HETEROGEN* OR PARTICIPATION OR "SELF CONTAINED CLASSES" OR "LEAST RESTRICTIVE ENVIRONMENT" OR "LRE" OR "COGNITIVE DEVELOPMENT" OR "INTELLECTUAL DEVELOPMENT") AND (EMPIRI* OR QUANTI* OR QUALITATIVE OR MIXED-METHOD* OR EVIDENCE-BASED) NOT (UNIVERSIT* OR TERTIARY).

The initial search for Review III led to a corpus of 5417 records after duplicates were removed. These were retrieved from the databases Eric (N= 4356), Education Source (N= 529) and PsycInfo (N= 531) (access via EBSCO). After a first manual screening of the abstracts 407 studies met the inclusion criteria and a discussion of critical incidents (four-eye principle) led to a reduction of the corpus to 250 studies. Having screened the full texts, 150 studies could be included. In the end, two more duplicates were removed manually.

Results of Review III

All details of review III cannot be fully unfolded at this point, but the main results shall be given[3], before the results of the aggregation of all three reviews will be presented afterwards. Since a table of 148 studies is not easy to interpret and should therefore be visualized in a condensed form. Such a summary is hardly possible in human visualization. For this reason, a first glance of the relationship between studies shall be portrayed with the help of VOSviewer.

Figure 1. *Overview by VOSviewer.*

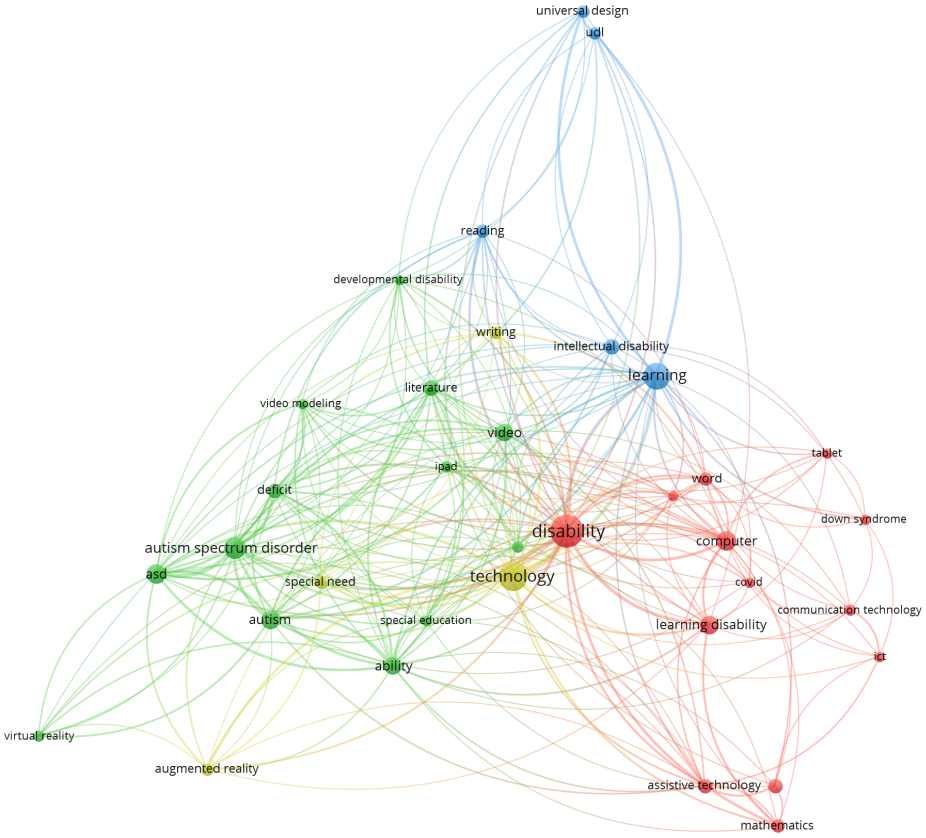


The graphic in figure 1 visualizes key terms co-occurring in research focused on special educational needs and digital media, drawn from the identified 148 studies. The nodes represent the frequency of individual terms in the analyzed studies. The larger the node, the more frequently the term is mentioned. The lines between nodes symbolize associations and correlations between terms, illustrating which concepts are often discussed together in the texts. Terms were selected based on their relevance and reduced to common synonyms or alternate spellings. For example, terms like "ASD," "autism," and "autism spectrum disorder" were grouped under the most frequently used term. In summary, figure 1 gives a first hint to the overall conclusion on research question one, namely that a) on the level of *SEN areas* autism spectrum disorder – followed by intellectual disabilities - was the most frequently analyzed area, that b) on the level of *media use* that computer was the most frequently analyzed tool and that c) on the level of *skills* reading and writing were most frequently analyzed.

[3] For a detailed overview of review III with a summary of the key findings of the respective studies you are referred to the table in the annex (along with the list of references)

Connections between terms indicate which concepts often appear in combination and thus show their mutual relationships. The colors highlight different thematic clusters. For instance, the term "ASD" (autism spectrum disorder) was found to be closely associated with concepts such as "virtual reality" and "video modeling", suggesting that these concepts are often discussed together in the scientific studies. Looking at the other nodes, a tentative interpretation is that in the context of visual impairment general terms like mobile technology and communication technology are core concepts prioritized in research.

Figure 2. *Summary of co-occurring key terms by VOSViewer.*



The second more detailed graphic (= figure 2) visualizes key terms co-occurring with terms from the area of "autism spectrum disorder" (=green node), "learning" (=blue node) and other forms of "disability" (=red node). The nodes in figure 2 show a more differentiated picture of frequently occurring terms without grouping synonyms or alternate spellings under one central term. Therefore, the graphic brings specific terms to the forefront that are particularly relevant to more specific aspects of e.g. "autism-related research", namely "asd", "autism", etc. The different colors again mark thematic groups that often appear together. For example, in case of the green node "autism spectrum disorder" there is a cluster with terms like "video modeling", "virtual reality" and "augmented reality" which suggests that these technologies are frequently researched in connection with "autism spectrum disorder". Another cluster linked to "autism" includes terms like "deficit," and "developmental disability" indicating that autism is often considered alongside other special educational needs. A tentative interpretation of the other nodes (e.g. the blue node) could indicate that the concept of "UDL" is highly relevant in the context of "learning" and – especially in the context of "reading". "Mathematics" (red node), on the contrary, seems to be particularly interconnected with "assistive technology".
All in all, it can be summed up that the results of review III are in line with the findings of reviews I and II. Particularities of review III refer to the fact that in the international review "autism" was addressed

more frequently in the corpus. In addition, the "UDL" was addressed more often than in reviews I and II. A similarity of review III with the previous ones was that "learning *with* media" was predominant whereas "learning *about* media" (in the sense of "media education") was neglected as a point of interest.

Aggregation of the results of the three reviews

Now let us take a look at the aggregated results of all three reviews. All three reviews followed the PRISMA criteria (Moher et al., 2009) and Cochrane guidelines (Higgins et al., 2019). In total, 34 German-language and 148 international studies from 2010 to 2023 were included, all with a student-centered perspective.

Similarities: All three reviews focused mainly on individual support for compensating for deficits, with less attention to cooperative learning within the classroom on shared topics. A significant emphasis was placed on the evaluation of training programs designed to provide individualized support for students. This indicates that the research focus in studies on the use of digital media for students with special educational needs tends to be on compensatory use of digital tools. Consequently, there is a clear research gap concerning studies on learning about media or on the acquisition of media literacy in the context of special educational needs and inclusion. The use of assistive technologies and strategies to improve accessibility was addressed less frequently than the use of digital media to acquire content knowledge in the field of maths, reading and writing.
Bosse et al. (2019) distinguish three levels on which digital media can help people to participate in society:

- The level "participation *in* media" refers to an equal representation of all humans in medial representations.
- The level "participation *at* media" – prevalent in studies from the international context – implies that everybody can have full access to content, whether through the use of assistive technology or due to the fact that the information is offered giving respect to the rules of the UDL (CAST, 2024) – offering *multiple means of engagement, multiple means of representation* and *multiple means of action and expression.* These guidelines have the overall aim of concrete suggestions that can be applied to any discipline or domain to ensure that all learners can access and participate in "meaningful, challenging learning opportunities" (CAST, 2024).
- The last dimension, participation *via* media, refers to media competence in a broader sense and comprises both the level of "learning *about* media" and the more instrumental aspect of "learning *with* media". In terms of participation *in, at*, and *via* (Bosse et al. 2019) media, the focus of most empirical studies in all three reviews was primarily on participation *via* media. Studies investigating the use of digital media for participation *in* media were underrepresented.

Regarding the learning and support areas addressed, most studies focused on performance deficits in reading, mathematics and writing. Only a few studies formulated guiding research questions independent of specific subject areas, aiming instead for broader goals, such as the design of inclusive teaching environments. In terms of the SEN-area, most studies addressed the area of "learning," followed closely by "autism" (especially in the international studies). As a research desideratum, it is suggested that areas within the spectrum of special educational needs (SEN) beyond "learning" and "autism" should receive greater attention from the research community.

Differences: Notably, both German-language reviews predominantly employed quantitative research methods, whereas the international studies displayed a relatively balanced mix of quantitative, qualitative, and mixed-methods approaches. More studies on participation *at* media were found in the international context, particularly concerning Universal Design for Learning (www.cast.org) (=review III).
As can be seen from the graph below (figure 3) the setting of the majority of studies is secondary education – both in the German reviews (I & II) as in the international review (III). Special needs schools are a typical German phenomenon and this cultural difference might explain the high number of SEN

school contexts in reviews I and II compared to the international context. As already elaborated in the detailed comparative synthesis of the international results the dominant areas of research are "learning disorders" and "autism spectrum disorders" (figure 4). With regard to this finding, there is a parallel between both the German reviews and the international review.

Figure 3. *School types*.

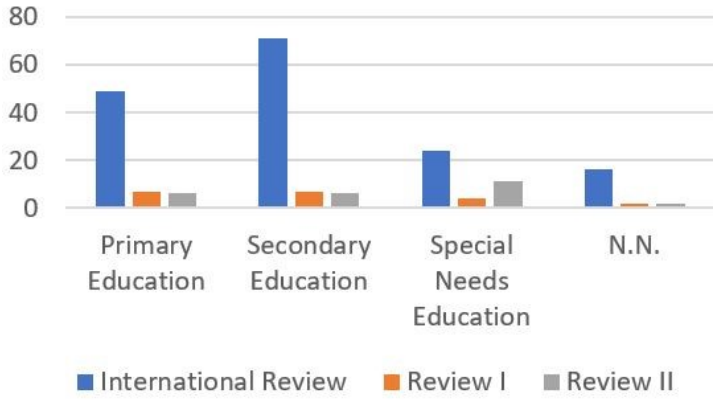
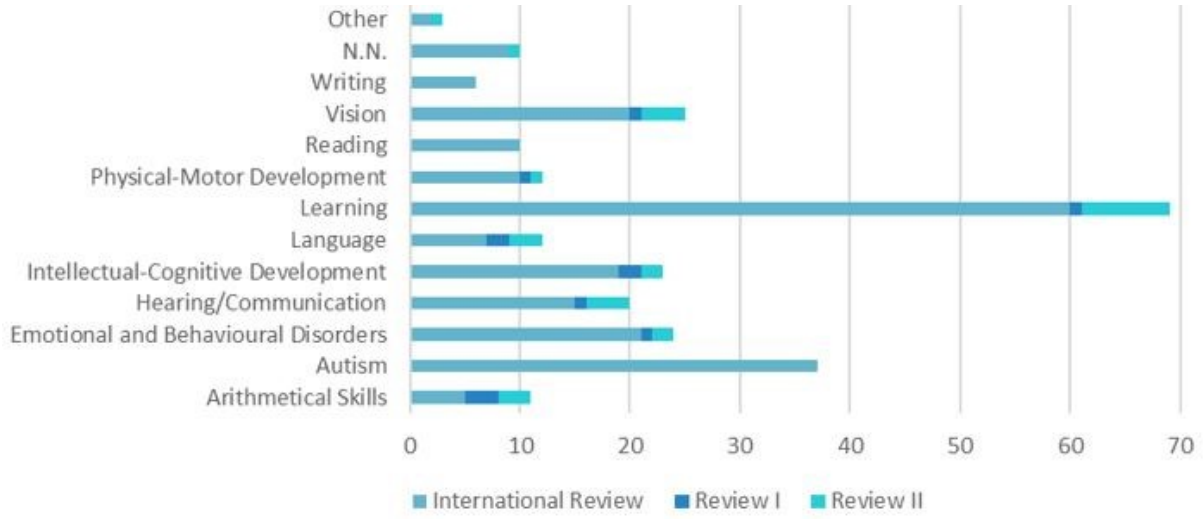


Figure 4. *Types of SEN.*



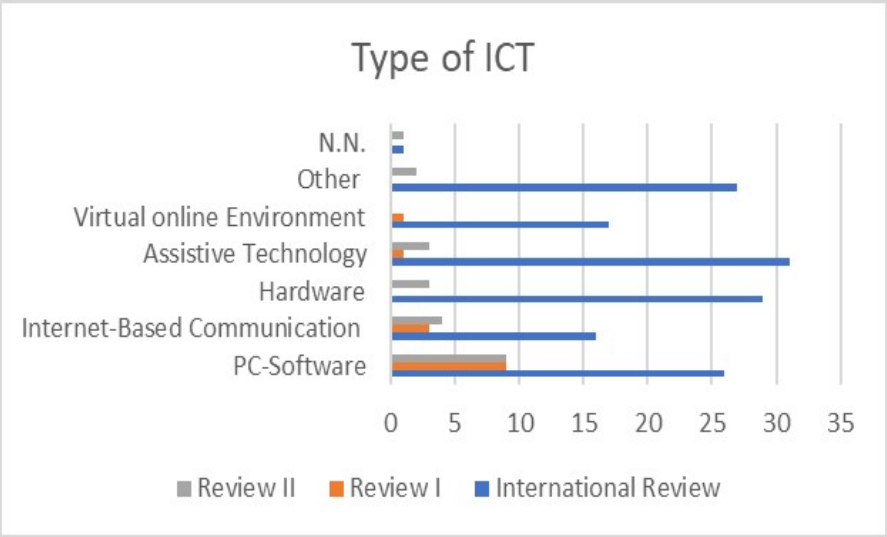
Looking at the type of settings (inclusive, integrative vs. excluding), it is encouraging to see that the total number of studies in the inclusive and integrative settings sums up to 86 studies with only 42 studies from excluding settings. This indicates that the Convention on the Rights of Persons with Disabilities (CRPD) – guaranteeing everyone's right to an education that provides equal and fair opportunities for all – is beginning to show results.

For further research it is important to note that most of the studies are didactic-oriented in the sense that media are analyzed in their "tool"-function. It has been a recurring result of our research in the domain that "participation *via* media" is over-represented as a research item in most of the studies (with a clear predominance on learning *with* media - neglecting more or less the acquisition of a critical media competence (= learning *about* media).

The last graph refers to the type of ICT which is the focus of the research. Adding the numbers identified in the three reviews we can see the strongest focus on PC-Software (N=44), which goes along with the previous finding (enabling participation via media). The second highest figure refers to assistive technology (N=35), followed by research on hardware (N= 32). The smallest number of studies deals with

the analysis of a virtual/online environment (N=18) – which can be due to the fact that the technology is more recent.

Figure 5. *Types of ICT.*



Summary and discussion of the aggregated findings

Like in the analysis of the previous publications by the authors, the strongest statement that can be drawn from the aggregated research is that all in all, digital optimism seems to be justified. Most of the studies report positive findings. The evidence from the studies suggests that most empirical studies concentrate on how to use digital media to learn effectively. Comparing German studies (I & II) to those from the international context (III), we can conclude that in the English studies there is a slightly bigger emphasis on the Universal Design for Learning and on assistive technology. As far as the areas of SEN are concerned, there is a predominance of studies on learning difficulties and autism spectrum disorder. A strong desideratum emerging from both German research and from the international research can be seen in the fact that "learning *about* media" seems to be neglected (see above). Another suggestion for future research in the community can be derived from the finding that up to now most of the empirical studies stem from a quantitative research paradigm, whilst research from the qualitative paradigm is largely absent.

Methodological discussion: AI-driven screening

In the next section of this paper the methodological discussion moves on to a comparison of the manual screening process to the AI-driven screening processes. We shall now concentrate on benefits and flaws of the AI.

Characteristics and functionality of AI in the context of reviews: general pros and cons of AI

Before going into detail of the analysis, the functionality of the chosen automatic-screening software "ASReview" will briefly be described, as the choice of a trustworthy tool is delicate. After scientific analysis, Campos et al. (2024) have given the best score to the above-mentioned software for the aim of "screening smarter, not harder". Some of the advantages of the software are that it is open source and

based on a local or server installation with full control over the data. According to the - marketing-oriented - website of the ASReview software it follows the "Reproducibility and Data Storage Checklist for AI-Aided Systematic Reviews" (ASReview, 2025). The functioning of the app is algorithm-based: The entire dataset of identified abstracts meeting the search string must be loaded into the tool, and the AI is initially "trained" with "relevant" articles based on human decisions. Studies are marked as relevant or irrelevant based on the predefined screening criteria. Then the AI generates a descending order of articles based on relevance, which is manually reviewed until a "saturation" point is reached (the AI continuously updates the ranking based on the clicks and "learns" from this interaction). The review process is complete when "saturation" is achieved. Yet, it remains a human decision to decide when exactly saturation occurs. A possible criterion for saturation could be, for instance, when a high proportion of articles are classified as "irrelevant" by the AI in a sequence, i.e., when a so-called "plateau sequence" of irrelevant articles is reached:

> Screening prioritization re-arranges the records to be screened from random to a more intelligent order. Active learning denotes the scenario in which the reviewer is labeling specific records that are selected by a machine learning model. The machine learning model learns from the reviewers' decisions and uses this knowledge to select the next record presented to the reviewer. […] Active learning for systematic reviewing is very effective for systematic reviewing and can save up to 95% of screening time. (Rens, 2022)

In the present study, the dataset containing 5417 records was fed with 2 relevant and 2 irrelevant records as prior knowledge. Then, the software extracted relevant features on a more abstract level with the aim of reducing the data. "An active learning model consists of a feature extraction technique, a classifier, a query strategy, and a balance strategy. The default setup (TF-IDF, Naive Bayes, Maximum, Dynamic resampling) overall has fast and excellent performance" (ASReview, 2025). The feature extraction technique used for our review was "TF-IDF", the classifier "naive bayes", the query strategy "maximum" and the Balance strategy "dynamic resampling (double)". The AI iteratively classified the remaining data, refining its predictions based on continuous human input. This is called "Researcher-in-the-Loop". The advantage of the ASReview program is tremendous as far as saving time is concerned.

Nevertheless, several risks became evident. Even on condition that information retrieval and deduplication of the corpus are done via a classical database search (because query formulation with LLMs can lead to a bibliography with non-existing references) further challenges of AI screening stem, for example, from grey literature discovery. Tools often underperform on non-standard designs, qualitative studies, non-English abstracts, and grey literature. we cannot know which studies are missed. Although the point is irrelevant for the analysis at hand because grey literature was excluded anyway, the concern has to be taken seriously.

Apart from this, problems reside in the field of screening and prioritization. The automated active learning procedure to rank abstracts is kind of a "black box" for the researcher - who, in turn, is unable to grasp according to which criteria the ranking took place. For automatic screening processes you do not get the trails (who/what excluded what, when, and why). Automation may "hide" judgments (e.g. inclusion criteria interpretation, outcome hierarchies), and underlying procedures are subject to general AI bias. Training data mirror publication and indexing biases (English-language, high-income settings), risking systematic exclusion of under-represented studies. The language aspect is less important in the context of the review III because only English-language publications were intended and income plays a minor role for the research question, but the risk has to be kept in mind for other review contexts. If a review is conducted automatically it is supposed to be "objective"; but machine learning is also due to error.

Opaque training data and silent updates hinder replication; results may drift as models change. Therefore, reproducibility is questionable, AI-generated queries can be unstable across runs, as databases change over time; web sources and preprints are especially volatile. Moreover, performance varies by topic, year, and study design, and LLMs can misclassify study designs; but this is also true for human screening.

Another area of concern results from the fact that active learning can amplify initial seed-set choices; poor seed diversity leads to concept drift and missed studies. This is a serious limitation of the analysis at hand because taking two random studies that fulfill the criteria and another two that do not fulfill the criteria is risky in the sense that a bad choice has high consequences for deep learning.
Finally, a general problem area of systematic reviews (regardless of whether automated or not) stems from the fact that optimizing the research string for the sake of precision can reduce recalls —which is unacceptable if near-complete retrieval is essential.
To sum up the question on how far automation might replace dual independent processes, we argue that AI should prioritize or suggest, but not make final exclude/include decisions without human verification. A recommendation therefore is to double-check AI-extracted data. Apart from this, model and tool names as well as the version should be recorded: date of runs, seed studies, export files, exact queries; etc. AI can support the process of synthesizing data and automate parts of a screening process but it can never replace the expertise of human researchers. Overreliance risks de-skilling and undermining critical appraisal.
At the end of this pro and con argumentation it remains still an open question what loss in recall is acceptable for timeliness in living reviews. All in all, AI can safely reduce workload when used as assistance under strong human oversight, with validation and full transparency.
Further limitations and problem areas:

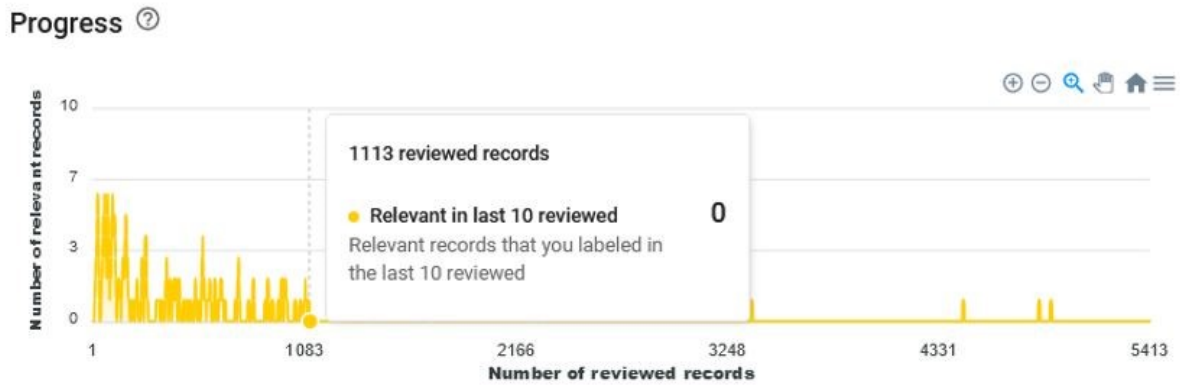
- **AI's susceptibility to error due to manual "irrelevant" or "relevant" clicks**: In case of wrong labeling in the training phase (e.g. due to human slips) the entire AI's ranking logic is affected, which can lead to skewed results.
- **Uncertainty about when "saturation" is reached**: A risk can be seen in the fact that one cannot be sure to not have overlooked valuable results when stopping the screening process. There are no objective criteria for deciding when to stop.
- **Abstracts in the corpus**: AI works using information (presumably keywords) from abstracts, whereas humans work with the full texts. A challenge for the AI, therefore, is that the abstracts do not necessarily reflect all of the predefined criteria in the search string. Abstracts may be incomplete regarding "universal" information. A concrete example can be seen in the fact that sometimes so-called hyperonyms are used in the abstract, like the generic term "digital media" instead of "VR/AR". The problem can also occur vice versa, namely when the abstract only contains hyponyms. A hyponym is a word whose semantics is more specific than its corresponding hyperonym, like "AR" instead of "digital media". The semantic field of a hyperonym, also known as a superordinate, is broader than that of a hyponym. The biggest challenge may occur in so-called "co-hyponoyms" when there are coordinate terms on the same hierarchical level. Yet, the complexity of the search string is limited, in the sense that the number of keywords is not infinite. Consequently, a deliberate decision FOR or AGAINST a special phrasing has to be made by the research team. Considering the area of Special Educational Needs, it might for example be the case that special types of "hearing disorders" like the co-hyponyms "conductive hearing loss"," schwannoma" or "meningioma" are not considered in a search string, whilst the hypernym "hearing disorder" is included. From an ethical perspective it is highly questionable to exclude one hyponym and to include another (just for frequency reasons). However, since the string is limited, certain arbitrary decisions must be made. Machine learning is imperatively based on statistics, so that e.g. frequent disorders like "attention deficit hyperactivity disorder" or "dyslexia" may deliberately be included, whereas a less frequent subtypes of hearing disorders may be omitted – knowing this can have discriminatory effects on the results.
- **Human errors whilst using the software**: Another problem may reside in the fact that human errors, like misprints in the title, may lead to flaws when working with ICT. A limitation of the software is that a simple search via "authors" and subsequent correction of misprints is not possible and a deviation in the title like using an abbreviation may lead to misinterpretations and mistakes.

All in all, AI can be expected to be more rigid and consistent in making decisions, thereby avoiding human bias. Nevertheless, it is an open question whether this is always an advantage or, on the opposite, more of a disadvantage because sometimes the decision on inclusion and exclusion demands sensitivity and contextual understanding. For instance, if one predefined inclusion criterion is research from the "perspective of pupils", it may be unclear what action is appropriate when a study focuses on the perspective of teachers but includes *one* instance of a pupil's perspective.

Comparison of our analysis to the AI-driven screening process

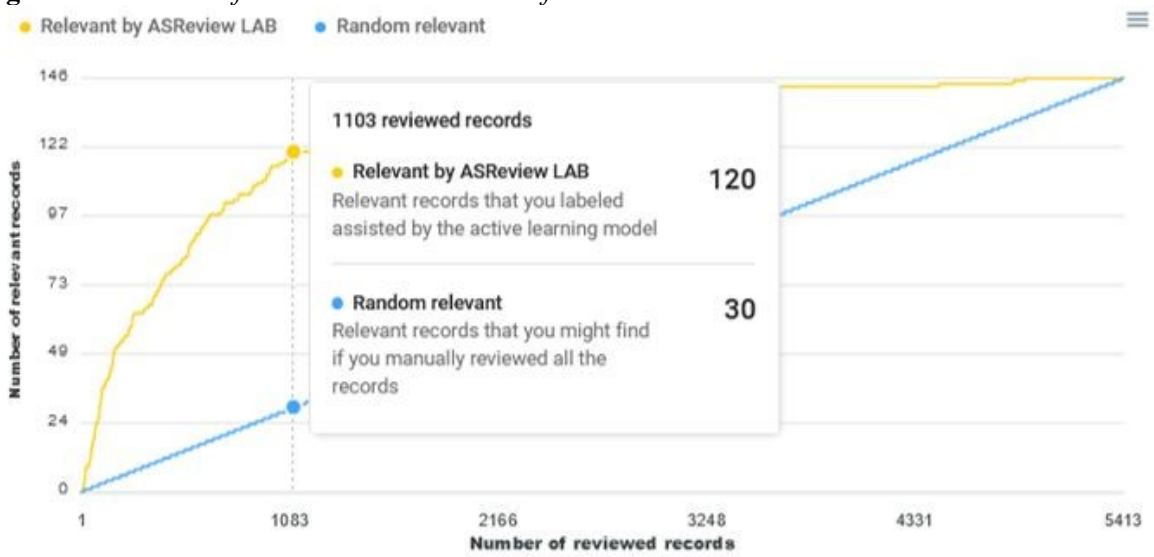
Having completed the manual screening process as well as the AI-driven screening process we were curious to cross-check where differences arose. This analysis highlights both the strengths and limitations of the AI tool, providing insights into its performance and potential challenges. To do so we generated an Excel file showing the "RelevantByLabellingOrder" with the help of the IT department, notably with the help of Jenny Paxian and the ICT-team. They designed an algorithm to retrieve the studies in the order of relevance. Having thus retrieved the included studies (N=148) in the order of the labelling history we could check which studies had been classified as the least relevant ones by the AI. Possibly, the last 24 manually-identified studies would have been skipped in a completely data-driven procedure, because after 124 relevant articles identified by the AI (after having screened 1113 abstracts in the original corpus, see figure 6) there was a "plateau" of about 100 articles in a sequence being classified as "irrelevant" by the AI – which might have led to stopping the procedure without knowing from the manual process that more relevant studies were about to pop up in case of going on.

Figure 6. *Screenshot from ASReview showing the progress.*



On the one hand, using the AI *only*, valuable insights might have been missed (namely the last 24 relevant studies), but on the other hand the AI-driven screening process is much smarter because with the AI-driven process you get almost 87% of the possible total results after screening just 1113 out of the total of 5417 studies in the corpus (which corresponds to a percentage of 20%). In other words, the input-output relation of the AI-driven screening turned out to be fine in the retrospective analysis. After having screened 1103 articles 120 had been classified as relevant by the ASReview LAB (Fig. 7).

Figure 7. *Screenshot from ASReview: number of viewed records.*



Nevertheless, we were curious to know *which* of the 148 studies in the dataset would have been missed exactly. Therefore, as a retrospective reflection of the process, we planned an analysis of the last 24 studies in order to deduce possible reasons for the late show-up of the cases of doubt.

When analyzing these with regard to the potential reason for "mis-classification" it gets evident that the most frequent reason for the late show-up seems to be a setting beyond school (126, 128, 129, 142) or that a hyponym (subordinate) is used in the abstract instead of hyperonym (superordinate) or vice versa (127, 135, 136, 144). Another typical source of error seems to be that the perspective of the study is different from pupils (130, 131, 143), that there is no SEN context (132, 147), that a misleading keyword is used in the abstract e.g. "review", "addiction" (138, 140), that the criterion "digital medium" is not fulfilled (141, 145) or that the language is other than English (148, 149). Once there is a methodological reason for exclusion, namely that it is a single case study (137); and once that the criterion "peer review" is not adequately fulfilled (146).

In Table 1 the potential deviance from the human process shall be run through one by one – although the reasoning for the possible late identification remains speculative.

Table 1. *Detailed description of possible exclusion criteria in the AI-driven screening process*

Study	Summary
126 (Byiers et al., 2014)	The setting was partly beyond formal school education. One of the three test persons was aged 15 (and thus a pupil) whilst the others were up to 47 years. Therefore, the criterion "population" only fitted partly.
127 (Bouck et al., 2017)	The term "digital media" was not mentioned in the abstract but just the hyponym "VR/AR". In addition, there was no focus on a special area of SEN but just a focus on fractional arithmetic.
128 (Flynn et al., 2023)	Just one video of one pupil was analyzed, so the setting "school" was not fully complied with.
129 (Saito et al., 2017)	This study turned out to be quite similar to the previous one – which led to the assumption that the AI might have learned from the re-labelling of 128.
130 (Awasthi et al., 2021)	The school-setting was addressed only indirectly because parents were supposed to learn how to support pupils. This study had already been marked as a borderline case in the manual screening as well.
131 (Lindsay et al., 2019)	At first sight the setting seemed to be a hospital, like in the previous sample. Nevertheless, the purpose of the MINT intervention with Lego bricks was to raise the competence of pupils; and pupils were interviewed. Like be-fore, this case seems to be a self-learning mechanism of the AI to include this study after 130 had been added manually.
132 (Kesler et al., 2016)	There was just a marginal reference to pupils with SEN: no special type of medium was addressed but the importance of critical media literacy was stressed.
133 (Istenič Starčič et al., 2013)	The focus is on understanding what a tangible user interface can add to teaching and learning (concept development). It was a design-based research study analyzing how the tangible user interface can support pupils with low motor skills. Perhaps this was a borderline case for the AI due to the non-experimental approach of the study. Yet, this is speculative.
134 (Wiest et al., 2020)	The authors can show that computerized cognitive training (CCT) improves the working memory of the participants. This is a highly interesting item because at first sight all the predefined inclusion criteria seem to be met. The reasons for inclusion/exclusion remain intransparent.
135 (Şenel & Kutlu, 2017)	This study deals with the SEN of students who are visually impaired: computerized adaptive testing (CAT) and reader-assisted paper-pencil testing (raPPT) are compared and student views about them are analyzed. Here the problem may reside in the use of the keywords. The abbreviations are perhaps too specific to be found via the AI tool.
136 (Bargagna et al., 2019)	This study came to conclude that robotic kits (bee-bot training) can help children with down syndrome. Results indicate that bee-bot robots are a very significant device for promoting interest, attention and interaction. Here it is un-clear why the AI software did not suggest the study earlier because all inclusion criteria seem to be met. Again, a plausible reason for the late identification might be that the key words like "bee-bot" might have been too specific. Yet, the decision is opaque.
137 (Vassilopoulou & Mavrikaki, 2016)	This study deals with attention deficit hyperactivity disorder and reports that a pupil profits from an ICT driven biology class. Here the late identification might be in connection with the fact that it is just a single case study.
138 (Mandanici et al., 2018)	This study fulfils all the inclusion criteria. It is a classical intervention study: Authors present technology-enhanced learning activities with the employment of three experimental applications. Yet, the reason for the late pop-up in the AI driven sequence might be that the word "review" might have led to the misin-terpretation that the paper was not an empirical study.
139 (Marchetti & Valente, 2016)	This study aimed at assessing the benefits of game-based learning for pupils affected by autism spectrum disorders. Apart from the lack of a general keyword such as "digital media", it is surprising that the study was found at such a late stage. This is another example of a shadowy decision of the AI.
140 (Fidan et al., 2021)	This study might be special in that it deals with addiction, mood disorders, narcissistic acts, and ethics violations. The relation to the predefined SEN areas is not that obvious because the focus is on adolescents' perceptions toward Instagram as a social network site. The findings, however, are highly relevant.

	Here the AI seems to be more "consequent" in the way that studies without the predefined catchwords are eliminated. It is a question of human bias if "addiction" is seen as an area of social-emotional development or not. The decision of the AI is logical in the sense that this study had been a borderline case in the human screening process, too.
141 (Chan et al., 2022)	The late appearance of the study seems to be due to a misinterpretation by the AI, because according to human check all the inclusion criteria are met: Students who are deaf and hard of hearing are in the focus with regard to their language acquisition skills fostered by gamification. Here the problem might reside in the fact that it is unclear whether the study is on digital games or not.
142 (Martínez-Álvarez, 2017)	This study deals with two first graders with learning difficulties. Yet, it is an after school-program and thus not directly in the center of the re-search interest. The children use a digital camera to document science and engage in multigenerational learning. Strictly speaking, the study only indirectly fulfils the in-clusion criteria because the population are pupils, and the setting is "schools", but it is informal learning not formal learning.
143 (Charnock & Standen, 2013)	This study analyses the development of a gendered identity for men with intellectual disabilities (contemporary self-understanding). Digital media are involved in the sense that games are seen as permissive spaces to try out ideas.
144 (Ölmefors & Scheffel, 2021)	This study is another example where all the inclusion criteria are perfectly met (neurodiversity and flipped classroom, setting: upper secondary school). For this study, no apparent reason for a late identification can be made out. Yet once again it is striking that the hyperonym "digital media" does not appear in the abstract, which might have been the reason for the late ranking by the AI.
145 (Cramér-Wolrath, 2013)	This study is a case of doubt due to the fact that a cochlear implant can be classified as an assistive technology and thus as a digital medium. On the other hand, it can also fall out of the classification "digital medium" if the focus is laid on the character as a medical tool. Therefore, this study is a case of doubt both for the AI-driven screening and for the human screening.
146 (Fenty & Allio, 2017)	This study is special in the sense that the basis of this journal article is a doctoral thesis, which - if it had not not published in a journal as well - would have been excluded. Apart from this it is not prototypical in the sense that it deals with the population between school and college. Last but not least, the SEN area stems from a broad definition of inclusion because it says "students from underrepresented populations". Again, the study is a case of doubt both for machine screening as well as for human screening.
147 (Lam et al., 2020)	This study is special in the way that from the 79 classes which are part of the analyzed population only two schools are special educational needs schools. The question arises whether it would have to be excluded because the majority of the population can be supposed to be neurotypic or whether even one single person in the sample leads to including the study. Again, we see AI-screening and human screening in accordance.
148 (Alananbeh & Asha, 2023)	This study deals with assistive technology again – this time in the context of visual impairment. For the AI it might have been misleading that the article stems from Jordan and has a second foreign abstract. A formalistic approach necessarily leads to exclusion if the inclusion criterion was "published in English". Interestingly, the AI driven decision would have been wrong because the main part of the text is in English and therefore meets the inclusion criteria.
149 (da Silva et al., 2018)	The same applies to the last study which has an English abstract as well as a Portuguise abstract: It is available in two languages. If the AI operates mechanically this might have led to the exclusion of these last two articles.

These findings are in line with a general caution in view of automatic screening as elaborated above, that is tools often underperform on non-standard designs, qualitative studies, non-English abstracts, and grey literature.

Conclusion

Let us come back to the initially mentioned twofold concern of this paper. The first aim was to aggregate the results of the three reviews.

(1) For the question of aggregating the results of the three papers, some general conclusions seem to be appropriate, namely that
   - **Digital optimism seems to be justified**: The aggregated findings suggest that digital tools can significantly support inclusive education, fostering positive outcomes for students with special educational needs (SEN).
   - **More studies are needed on "learning *about* media" (-> how to acquire media competence)**: Research into how students acquire media competence, particularly in inclusive settings, remains underrepresented and should be prioritized in future investigations.
   - **More research is needed from a qualitative research paradigm**: Most studies reviewed employed quantitative methods, leaving gaps in the deeper understanding of individual experiences and contextual nuances that qualitative research could uncover.

(2) For the second (methodological) question of the paper on whether to recommend AI-driven screening processes or not, the following interim conclusion can be drawn:
   - **Efficiency of AI-driven screening:** While the AI-driven process did not achieve the claimed 95% time-savings, it still significantly reduced the workload by identifying approximately 87% of the relevant results after reviewing only 20% of the dataset.
   - **Quality of results:** The AI-driven screening successfully identified most relevant studies, but the exclusion of 24 manually included studies highlights its limitations, particularly in handling ambiguous or complex inclusion criteria.
   - **Dependence on human oversight:** The AI process demonstrated that human input remains critical for resolving edge cases, interpreting contextual nuances, and mitigating the potential for biases in keyword-based screening.

Practical implications for teachers, school leaders and policy makers

Teachers, school leaders and policymakers can use the (all in all) positive and optimistic findings of the reviews when conceiving didactic learning arrangements which allow pupils with SEN in inclusive settings to participate *at* and *via* media (see above Bosse et al., 2019). As all three reviews show, digital media turn out to be conducive to learning – both as far as the acquisition of basic reading, writing and calculating skills is concerned and as far as the acquisition of social skills is concerned. Especially for pupils with autism spectrum disorder digital media are a great benefit.

For school leaders this means that spending money on digital media and software for their institution is worth the investment; there should be (further) education for (future) teachers to familiarize them with the functionality of digital media, respectively the inclusive potential of digital media. It is important to invest in further education for the staff because media should not only be used by the teachers for the sake of making the preparation of classes easier but they should be used for the sake of the skills development of pupils. A typical win-win situation, where teachers save time and pupils are supported in their learning process. Yet, it has to be kept in mind that digital classes do not necessarily run by themselves. Some students need a lot of support here. This means that extra staff is required for the maintenance of the digital infrastructure.

For policy makers this means that the acquisition of media competence should be a mandatory part of school curricula. Finally, digital media learning settings should also be established beyond the formal school context because similar positive effects on participation in society can be expected in informal settings as well.

Last but not least, media should be used in a preventive function and not only in a compensating, deficit-oriented remedial function. This means that by using multiple means of engagement, of representation and of action and expression (as asked for in the Universal Design for Learning) media can be used for constructive knowledge acquisition.

Implications for researchers

a) There is a general research gap in the field of learning *about* media. A desideratum can be seen in the finding that learning *about* media, like e.g. the empirical research on training programs for pupils to make them *critically reflect medial AI-output*, etc., was underrepresented in the corpus. This leads to the practical implication of giving more energy to these matters in SEN settings. Future research is needed in this area.
b) AI-driven screening has pros and cons; to benefit from the automation process, human supervision is necessary. In summary, AI-driven screening has significant potential as a complementary tool in systematic reviews, offering efficiency in managing large datasets and expediting the review process. However, its integration requires a co-constructive approach that combines AI's strengths with human expertise to ensure both time savings and accurate results. Future iterations of AI tools should focus on refining training protocols, incorporating diverse datasets, and improving mechanisms to handle ambiguities in abstracts and keywords. Ethical considerations, such as addressing biases inherent in AI models that prioritize frequent over rare conditions, are critical for ensuring inclusivity. Moreover, the adoption of AI in systematic reviews necessitates cultural and institutional shifts, including researcher training and discussions on balancing automation with human judgment. The findings also emphasize the importance of oversight in detecting relevant studies that fall outside standard keyword parameters or involve nuanced research settings. While AI-driven screening provides valuable support, it should augment, not replace, human expertise, ensuring that final selection processes are guided by informed judgment to achieve an optimal balance between efficiency and accuracy.

Conflict of Interest

The Authors confirm that no conflict of interest exists.

References

Alananbeh, N., & Asha, E. (2023). Obstacles to using assistive technology for students with visual impairments in Jordan. *Dirasat: Educational Sciences, 50*(1), 92–101. https://doi.org/10.35516/edu.v50i1.4507

ASReview. (2025). *ASReview: Smarter systematic reviews with open-source AI*. https://asreview.nl

Awasthi, S., Aravamudhan, S., Jagdish, A., Joshi, B., Mukherjee, P., Kalkivaya, R., & Patel, D. (2021). Transitioning ABA services from in clinic to telehealth: Case study of an Indian organization's response to COVID-19 lockdown. *Behavior Analysis in Practice, 14*(4), 893–912. https://doi.org/10.1007/s40617-021-00600-9

Bargagna, S., Castro, E., Cecchi, F., Cioni, G., Dario, P., Dell'Omo, M., & Laschi, C. (2019). Educational robotics in Down syndrome: A feasibility study. *Tech Know Learn, 24*(2), 315–323. https://doi.org/10.1007/s10758-018-9366-z

Bosse, I., Haage, A., Kamin, A.-M., & Schluchter, J.-R. (2019). Medienbildung für alle: Medienbildung inklusiv gestalten. In M. Brüggemann, E. Sabine, & T. Angela (Eds.), Medienbildung für alle – Digitalisierung. Teilhabe. Vielfalt(pp. 207–219). kopaed. https://www.gmk-net.de/2018/09/20/medienbildung-fuer-alle-medienbildung-inklusiv-gestalten/

Bouck, E. C., Bassette, L., Shurr, J., Park, J., Kerr, J., & Whorley, A. (2017). Teaching equivalent fractions to secondary students with disabilities via the virtual–representational–abstract instructional sequence. *Journal of Special Education Technology, 32*(4), 220–231. https://doi.org/10.1177/0162643417727291

Byiers, B. J., Dimian, A., & Symons, F. J. (2014). Functional communication training in Rett syndrome: A preliminary study. *American Journal on Intellectual and Developmental Disabilities, 119*(4), 340–350. https://doi.org/10.1352/1944-7558-119.4.340

Campos, D. G., Fütterer, T., Gfrörer, T., Lavelle-Hill, R., Murayama, K., König, L., Hecht, M., Zitzmann, S., & Scherer, R. (2024). Screening smarter, not harder: A comparative analysis of machine learning screening algorithms and heuristic stopping criteria for systematic reviews in educational research. *Educational Psychology Review, 36*(1), Article 19. https://doi.org/10.1007/s10648-024-09862-5

CAST. (2024). *Universal Design for Learning Guidelines version 3.0*. https://udlguidelines.cast.org

Chan, G. L., Santally, M. I., & Whitehead, J. (2022). Gamification as technology enabler in SEN and DHH education. *Education and Information Technologies, 27*(7), 9031–9064. https://doi.org/10.1007/s10639-022-10984-y

Charnock, D., & Standen, P. J. (2013). Second-hand masculinity. *International Journal of Game-Based Learning, 3*(3), 43–53. https://doi.org/10.4018/ijgbl.2013070104

Cramér-Wolrath, E. (2013). Sequential bimodal bilingual acquisition: Mediation using a cochlear implant as a tool. *Deafness & Education International, 15*(4), 201–221. https://doi.org/10.1179/1557069X13Y.0000000023

da Silva, E. D., Teixeira, J. A. de M. e S., & Scharlach, R. C. (2018). Utilização de estéreos pessoais por alunos do ensino médio: Mensuração da intensidade utilizada e hábitos de uso. *CoDAS, 30*(5), e20170124. https://doi.org/10.1590/2317-1782/20182017124

Fenty, N. S., & Allio, A. (2017). Using distance learning to impact access of diverse learners to advanced placement programs. *The Quarterly Review of Distance Education*.

Fidan, M., Debbağ, M., & Fidan, B. (2021). Adolescents like Instagram! From secret dangers to an educational model by its use motives and features: An analysis of their mind maps. *Journal of Educational Technology Systems*. https://doi.org/10.1177/0047239520985176

Flynn, E., Motiff, M., Mueller, M. K., & Morris, K. N. (2023). Testing of the Facial Expression Coding System (FACES) for middle school–aged special education students and development of a training protocol. *International Journal of Behavioral Development*. https://doi.org/10.1177/01650254231167313

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley. https://doi.org/10.1002/9781119536604

Istenič Starčič, A., Cotič, M., & Zajc, M. (2013). Design-based research on the use of a tangible user interface for geometry teaching in an inclusive classroom. *British Journal of Educational Technology*. https://doi.org/10.1111/j.1467-8535.2012.01341.x

Kesler, T., Tinio, P. P. L., & Nolan, B. T. (2016). What's our position? A critical media literacy study of popular culture websites with eighth-grade special education students. *Reading & Writing Quarterly*. https://doi.org/10.1080/10573569.2013.857976

Lam, P., Lam, P. L. C., Ng, H. K. Y., Tse, A. H. H., et al. (2020). E-learning technology and the advancement of practical constructivist pedagogies: Illustrations from classroom observations. *Education and Information Technologies*. https://doi.org/10.1007/s10639-020-10245-w

Lindsay, S., Kolne, K., Oh, A., & Cagliostro, E. (2019). Children with disabilities engaging in STEM: Exploring how a group-based robotics program influences STEM activation. *Canadian Journal of Science, Mathematics and Technology Education, 19*(4), 387–397. https://doi.org/10.1007/s42330-019-00065-0

Mandanici, M., Altieri, F., Rodà, A., & Canazza, S. (2018). Inclusive sound and music serious games in a large-scale responsive environment. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12630

Marchetti, E., & Valente, A. (2016). What a tangible digital installation for museums can offer to autistic children and their teachers. *International Journal of Game-Based Learning, 6*(2), 20–36. https://doi.org/10.4018/ijgbl.2016040103

Martínez-Álvarez, P. (2017). Multigenerational learning for expanding the educational involvement of bilinguals experiencing academic difficulties. *Curriculum Inquiry*. https://doi.org/10.1080/03626784.2017.1324734

Mertens, C., Quenzer-Alfred, C., Kamin, A.-M., Homrighausen, T., & Mays, D. (2022). Empirischer Forschungsstand zu digitalen Medien im Schulunterricht in inklusiven und sonderpädagogischen Kontexten – eine systematische Übersichtsarbeit. *Zeitschrift für Empirische Sonderpädagogik*, 14, 26–46. https://doi.org/10.25656/01:25529

Mertens, C., Quenzer-Alfred, C., Kamin, A.-M., Homrighausen, T., & Mays, D. (n.d.). *Inclusion going digital: An international perspective on empirical studies. A mapping review on digital inclusion in schools*. Manuscript submitted for publication in *Journal of Research in Special Educational Needs*.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). *Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement*. *Annals of Internal Medicine, 151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Ölmefors, O., & Scheffel, J. (2021). High school student perspectives on flipped classroom learning. *Pedagogy, Culture & Society*. https://doi.org/10.1080/14681366.2021.1948444

Quenzer-Alfred, C., Mertens, C., Kamin, A.-M., Homrighausen, T., & Mays, D. (2023). Systematisches Review des empirischen Forschungsstands zu digitalen Medien für Schüler:innen mit einem zusätzlichen oder einem sonderpädagogischen Förderbedarf unter Berücksichtigung inklusiver, integrativer und exkludierender Unterrichtsszenarien. In K. Scheiter & I. Gogolin (Eds.), *Bildung für eine digitale Zukunft* (Edition ZfE, Vol. 15, pp. 125–158). Springer VS. https://doi.org/10.1007/978-3-658-37895-0_6

Rens. (2022, April 12). Active learning explained. *ASReview*. https://asreview.nl/blog/active-learning-explained/

Saito, H., Ando, A., Itagaki, S., Kawada, T., Davis, D., & Nagai, N. (2017). Development of a support application and a textbook for practicing facial expression detection for students with visual impairment. *International Association for Development of the Information Society*. https://research.ebsco.com/linkprocessor/plink?id=cbc203a2-0494-33a9-8085-e3d00c3d53fc

Şenel, S., & Kutlu, Ö. (2017). Comparison of two test methods for VIS: Paper-pencil test and CAT. *European Journal of Special Needs Education*. https://doi.org/10.1080/08856257.2017.1391014

Stalder, F. (2016). *Kultur der Digitalität*. Suhrkamp. https://content-select.com/de/portal/media/view/56d5637e-dd60-4a67-bf67-3b3eb0dd2d03?forceauth=1

Vassilopoulou, A., & Mavrikaki, E. (2016). Can ICT in biology courses improve AD/HD students' achievement? *Improving Schools, 19*(3), 246–257. https://doi.org/10.1177/1365480216647144

Wang, Z., Nayfeh, T., Tetzlaff, J., O'Blenis, P., & Murad, M. H. (2020). Error rates of human reviewers during abstract screening in systematic reviews. *PLOS ONE, 15*(1), e0227742. https://doi.org/10.1371/journal.pone.0227742

Wiest, D. J., Wong, E. H., Bacon, J. M., Rosales, K. P., & Wiest, G. M. (2020). The effectiveness of computerized cognitive training on working memory in a school setting. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3634