

Valide: An Augmented Intelligence Infrastructure for Large-Scale Regulatory Validation of Digital Textbooks

Andressa Carvalho^{1,2}, Ranilson Paiva^{1,2}, Janaina Xisto^{1,2},
Lael Nervis², Ricardo Moro², Helenita Soares^{2,3}, Naelson Douglas², Alan Pedro^{1,2}

¹ Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Maceió, AL – Brazil

² Núcleo de Excelência em Tecnologias Sociais (NEES)
Universidade Federal de Alagoas (UFAL) – Maceió, AL – Brazil

³ Universidade Estadual de Alagoas (UNEAL) – Alagoas – Brazil

{andressa.carvalho, ranilson.paiva, alan.pedro}@nees.ufal.br

Abstract. *This paper presents Valide, an augmented intelligence infrastructure developed to support the technical evaluation of digital textbooks in the Programa Nacional do Livro Didático – PNLD. The platform integrates deterministic, AI-based, and hybrid validators with human review, traceable evidence, and institutional governance. The empirical analysis reported in this paper focuses exclusively on deterministic validators used in two official PNLD evaluation cycles encompassing 1,706 digital volumes. Results showed confirmation rates above 95% for the findings generated by these validators after human review. The findings provide evidence about the operational behavior of deterministic automation within the broader augmented intelligence infrastructure implemented by Valide, but do not evaluate the performance of its AI-based or hybrid validators.*

1. Introduction

The digital transformation of education has expanded the use of interactive digital textbooks in public educational programs, driven by education modernization policies and the growing demand for accessible educational resources [UNESCO 2021]. In this context, digital books have assumed a strategic role in teaching and learning processes, expanding pedagogical possibilities, accessibility, and educational personalization, especially within digital learning ecosystems [Ozbay and Ugurelli 2023, Larson 2010]. However, the increasing technical complexity of these materials also introduces challenges related to regulatory compliance, digital accessibility, and quality assurance, particularly when evaluation occurs at scale [Paiva et al. 2025, Sobrinho et al. 2023].

In public programs for the acquisition and distribution of educational resources, the technical evaluation of these works constitutes a critical stage to ensure the quality, accessibility, and pedagogical usability of the educational resources distributed to public schools [Marques et al. 2023]. Traditionally, these processes are conducted through predominantly manual analyses, supported by fragmented and weakly integrated verification procedures. However, the increasing volume of submitted works and the growing complexity of evaluation criteria progressively make it difficult to sustain essentially human-centered workflows without compromising consistency, traceability, and reliability of results.

In response to this scenario, different approaches based on automation and AI have been applied to the validation of digital educational content [Sobrinho et al. 2023, Tenório et al. 2025]. Despite advances achieved in automating specific technical verifications, challenges remain regarding the operationalization of these approaches in large-scale regulatory environments, especially concerning transparency, auditability, and structured integration with human decision-making processes [Wirtz et al. 2019].

Although the literature presents tools and models focused on technical validation and digital accessibility, there are still few systematic descriptions of how to architect and govern automated validation mechanisms that simultaneously integrate auditable requirements, rule-level traceability, and institutional human supervision. Furthermore, empirical evidence derived from continuous institutional operations capable of analyzing the stability, reliability, and behavior of automation under simultaneous increases in scale and regulatory complexity remains scarce.

Beyond technical and regulatory challenges, failures in digital textbooks may directly impact the educational experience of students and teachers. Structural problems, lack of accessibility resources, navigation inconsistencies, and compliance errors may compromise usability, content comprehension, and equitable access to digital educational resources, especially in large-scale public education contexts.

In this context, the augmented intelligence paradigm emerges as a particularly suitable approach for educational regulatory environments. In contrast to strategies oriented toward full automation, augmented intelligence emphasizes complementarity between human and computational capabilities, using Artificial Intelligence (AI) as a mechanism to support human judgment [Yang et al. 2021]. This perspective is especially relevant in scenarios where traceability, auditability, and institutional accountability constitute central requirements of the evaluation process.

The *Programa Nacional do Livro Didático - PNLD* (Brazilian National Textbook Program), one of the largest public textbook acquisition programs in the world, provides a relevant empirical context for examining these challenges. In recent years, the program has incorporated interactive digital textbooks and rigorous accessibility requirements, significantly increasing the technical complexity of evaluations [Marques et al. 2023]. This scenario highlights the need for solutions capable of supporting the technical quality, accessibility, and reliability of digital educational resources distributed at scale, ensuring better conditions of use for students, teachers, and public education systems. This article presents Valide, an augmented intelligence infrastructure developed to support the technical evaluation of digital textbooks in the context of the PNLD. The system integrates deterministic, AI-based, and hybrid validators with human review, traceable evidence, and institutional governance. Technical and regulatory requirements are operationalized through modular automated validators, designed as atomic verification units capable of producing traceable evidence associated with the evaluated rules. Their execution is controlled by versioned presets specific to different objects and stages of the evaluation process.

Beyond the technical architecture, this work proposes an adaptive governance model for automation, in which human validation is maintained as a central element of the decision-making process. The model establishes criteria for submitting findings to hu-

man review and defines conditions for the controlled transition between higher and lower automation regimes, preserving traceability, auditability, and institutional accountability.

The investigation is guided by the following research question: **How can an augmented intelligence infrastructure support auditable and scalable regulatory validation of digital textbooks while preserving human decision-making authority?**

To answer this question, Valide is analyzed as an empirical case implemented within a real public policy context. The analysis considers two complete official evaluation cycles and is restricted to deterministic validators based on explicit rules. Although AI-based and hybrid validators are part of the platform, their performance is outside the scope of this study. The goal is to examine the confirmation rates of automated findings across different requirement types and identify situations in which human review remains essential.

The main contributions of this work are:

- **(i) Artifact/Design:** a modular architecture for large-scale regulatory validation based on versioned atomic validators, with evidence and *logs* per rule, enabling reproducibility and auditability;
- **(ii) Method/Governance:** an adaptive automation governance model that explicitly separates automated detection from institutional decision-making, integrating human validation and evidence-based automation;
- **(iii) Evidence/Operation:** empirical evidence from two official PNLD Digital evaluation cycles, characterizing the confirmation rates of findings generated by deterministic validators in a large-scale supervised workflow.

Unlike previous studies focused on isolated technical validation tools, this work investigates the operationalization of augmented intelligence in a real large-scale educational regulatory environment under continuous institutional governance. Its contribution lies primarily in the orchestration, traceability, and governance of heterogeneous validation mechanisms rather than in evaluating the performance of AI algorithms. Although the study is grounded in the PNLD context, the architectural, governance, and operational principles presented are potentially applicable to other digital educational ecosystems that require technical evaluation, accessibility, traceability, and human supervision.

2. Context

The PNLD is one of the main Brazilian public policies aimed at the acquisition and distribution of textbooks and pedagogical materials for public education [Zambon and Terrazzan 2013]. The *Ministério da Educação - MEC* (Ministry of Education) defines the program guidelines and supervises the pedagogical evaluation, while the National *Fundo Nacional de Desenvolvimento da Educação - FNDE* (Fund for Educational Development) coordinates operational execution, including technical, regulatory, and logistical verifications.

The PNLD cycle involves interdependent stages that begin with the publication of the call for proposals and end with the distribution of the works to schools. As illustrated in Figure 1, this process may last approximately two years, involving multiple institutional actors, publishers, and evaluation teams.



Figura 1. PNLD cycle and main institutional stages.

During the cycle, the FNDE conducts verifications intended to ensure the conformity of the works with the technical and regulatory requirements established in the call for proposals, including structural, editorial, and digital accessibility attributes. These analyses involve hundreds of works and multiple technical criteria, requiring processes capable of operating with consistency, traceability, and scalability.

Historically, these verifications have predominantly depended on specialized human work. Although essential, this model presents challenges related to scale, operational overload, variability among evaluators, and deadline pressure. Strategies based solely on expanding teams or adding multiple layers of review tend to increase costs and operational complexity.

In this context, augmented intelligence approaches emerge as an alternative to support the evaluation process. In this paradigm, computational mechanisms perform repetitive and formalizable technical verifications, while human judgment remains responsible for contextual interpretation and institutional decision-making [Xu and Gao 2025]. The goal is to increase consistency, coverage, and traceability without replacing human supervision.

Valide operates in this context as a governable automated validation infrastructure developed to support PNLD stages that require systematic technical inspections and the production of traceable evidence. As illustrated in Figure 1, the system is mainly used during the stages of submission validation of the works (Stage 2), analysis of editorial attributes (Stage 5), and analysis of accessibility attributes (Stage 6), supporting the verification of technical and regulatory compliance of the works submitted to the program.

3. Related Work

Related work is situated at the intersection of technical validation of digital publications, accessibility, and assisted automation in educational regulatory contexts. Among the main tools used are *EPUBCheck*, aimed at the structural validation of EPUB files, and *ACE by DAISY*, focused on automated accessibility verification based on WCAG standards [DAISY Consortium 2024b, DAISY Consortium 2024a]. Solutions such as *AChecker* [Gay and Li 2010], *SortSite* [Abduganiev 2017], and *Total Validator* [Padure and Pribeanu 2020] are also widely used.

Although these tools perform relevant technical verifications, the literature highlights limitations of full automation in requirements that depend on contextual interpretation and human supervision [Vigo et al. 2013, Alsaedi 2020]. In this scenario, augmented intelligence approaches have been considered more suitable for regulatory environments that require traceability, auditability, and institutional governance [Yang et al. 2021, Wirtz et al. 2019].

Tabela 1. Comparison between validation tools and the Valide ecosystem

Criterion	EPUBCheck	ACE	AChecker	SortSite	Total Validator	Valide
EPUB structural validation	✓	–	–	–	–	✓
Automated accessibility checks	–	✓	✓	✓	✓	✓
Formalizable-rule validation	✓	✓	✓	✓	✓	✓
Context-dependent human review	–	–	–	–	–	✓
PNLD requirement mapping	–	–	–	–	–	✓
Versioning by PNLD object or stage	–	–	–	–	–	✓
Evidence recording	✓	✓	✓	✓	✓	✓
Rule-level metrics and history	–	–	–	–	–	✓
Institutional review workflow	–	–	–	–	–	✓
Batch execution	✓	✓	–	✓	✓	✓

✓ = supported; – = not identified as a primary feature.

Table 1 shows that established tools are mainly focused on specific structural or digital accessibility verifications. EPUBCheck, ACE, AChecker, SortSite, and Total Validator are included solely as comparative references and are not used as components of Valide. The technical contribution of Valide comprises a distinct modular architecture based on atomic validators, versioned presets, rule-level evidence, and institutional governance, integrating automated validation with supervised human decision-making in continuous large-scale operation. Previous PNLD studies described specific validation, triage, or digital-transformation initiatives, whereas this article systematizes the architecture and governance model implemented in Valide and evaluates its deterministic validation layer across two complete official cycles.

4. The Valide System

This section presents Valide, an augmented intelligence infrastructure developed to support the large-scale technical evaluation of digital textbooks in the context of the PNLD. The system integrates automated validation, human supervision, and institutional governance mechanisms focused on traceability, auditability, and regulatory control of the evaluation process. The conceptual foundations of the system, its validation model, operational architecture, and governance mechanisms that support its large-scale execution are also described.

4.1. Foundations and Validation Model

Technical evaluation is performed through automated validators, which constitute the fundamental units of the system. Each validator represents an atomic unit associated with a specific technical or regulatory requirement, responsible for verifying structural, technical, and digital accessibility criteria. Validators are version-controlled and linked to standardized messages and technical evidence generated during execution, ensuring traceability and longitudinal auditability.

Before validator execution, the system uses pre-validators to detect critical structural integrity failures, such as missing essential files or inconsistencies that would make subsequent stages unfeasible. This mechanism reduces error propagation and increases processing efficiency.

The conceptual model of Valide is based on four principles: (i) human centrality, (ii) modularity, (iii) auditability, and (iv) continuous evolution. Human centrality defines the system as support for institutional decision-making, with analysts responsible for validating or discarding findings whenever the requirement demands contextual

interpretation. Modularity organizes validators as independent components, facilitating maintenance and regulatory adaptation. Auditability associates each validation with explicit rules, controlled versions, and verifiable evidence. Finally, continuous evolution enables the incremental refinement of validators based on operational results.

Valide supports three validation strategies: (i) deterministic validators, based exclusively on explicit rules; (ii) AI-based validators, used in verifications dependent on computational inference; and (iii) hybrid validators, which combine both approaches. Adaptive automation governance is guided by operational evidence. Validators with stable behavior and high confirmation rates may evolve toward higher automation regimes, while requirements dependent on contextual interpretation remain subject to mandatory human validation. This model enables the gradual expansion of automation without compromising traceability, auditability, and institutional control.

4.2. Architecture and Operational Governance

The architecture of Valide was designed to support large-scale technical evaluations, prioritizing reliability, auditability, governance, and human supervision. The system organizes the validation process into independent modules responsible for automated execution of verifications, human analysis of findings, evidence production, and control of the regulatory workflow.

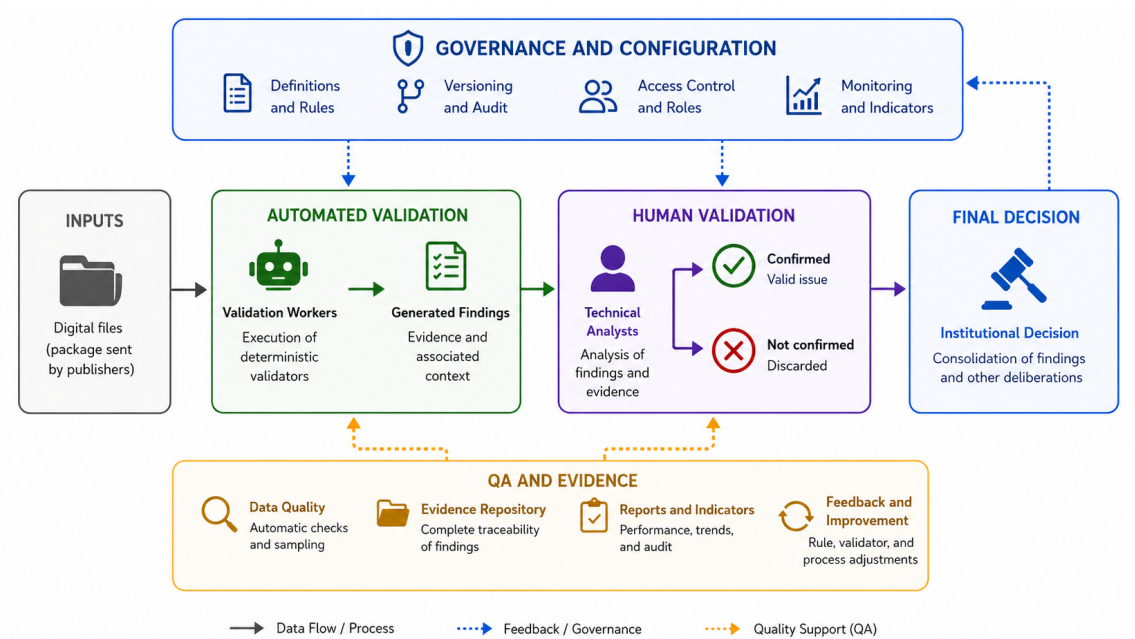


Figure 2. Conceptual architecture of Valide integrating automated validation, human supervision, and regulatory governance.

Figure 2 presents a conceptual overview of the system architecture. The workflow is organized into stages of automated validation, human validation, and institutional consolidation of decisions, integrated through governance mechanisms, Quality Assurance (QA), and production of traceable evidence. The validation workflow begins with the receipt of digital works submitted by publishers. Subsequently, automated validators

perform structural, technical, and digital accessibility verifications based on the presets configured for each object and stage of the evaluation process.

From an architectural perspective, Valide is organized into independent modules for structural pre-validation, validator execution, evidence persistence, human review, and governance. Pre-validators identify blocking failures before the remaining checks are executed, while the validation layer distributes independent validators across processing units according to the active preset. The infrastructure maintains persistent records of findings, evidence, execution logs, validator and preset versions, processing duration, and human-review outcomes. These records support monitoring, QA, reproducibility, and longitudinal analysis across evaluation cycles.

Presets constitute the main governance mechanism for the regulatory scope, explicitly defining which validators must be executed in each context. These configurations are versioned and recorded at the execution level, enabling reproducibility, comparison across cycles, and controlled adaptation to regulatory changes. As illustrated in Figure 2, the findings produced by automated validators are forwarded for human validation and subsequently consolidated into institutional decisions. The process is supported by continuous QA mechanisms, monitoring, evidence storage, and incremental refinement of validators.

Auditability is operationalized at the rule level. Each finding is associated with the corresponding regulatory requirement and recorded with minimum verifiable evidence, including identification of the affected file, problem location, and version of the executed validator. In addition, the system maintains institutional execution history, validator-level metrics, and QA records, enabling longitudinal tracking and review of results.

Table 2 summarizes the mapping between design requirements, implemented mechanisms, and evidence produced by the system.

Tabela 2. Traceability between design requirements, Valide mechanisms, and evidence produced in operation

Req.	Requirement	Mechanism	Evidence/artifact
R1	Rule-level auditability	Atomic validators with execution recorded by rule and version	Reports and logs with validator ID and version
R2	Regulatory traceability	Messages linked to the regulatory item with problem description	Regulatory identifier, description, and technical justification
R3	Minimum verifiable evidence	Minimum record for human inspection of the finding	File name/path and relevant line/segment
R4	Regulatory scope governance	Versioned presets by object and stage	Preset ID/version and change history
R5	Operational scalability	Modular execution and distributed validation processing	Execution, duration, and processing records of analyses
R6	Robustness against initial failures	Pre-validators for structural blocking	Records of critical failures and pipeline blocking
R7	Human centrality and accountability	QA and human validation with separation between detection and decision	Confirmed/non-confirmed classification
R8	Reproducible continuous evolution	Code versioning and execution-level recording	Commit/build references in logs and reports

5. Methods

This study adopts an artifact-centered Design Science Research perspective [Hevner et al. 2008] and conducts a longitudinal observational evaluation of Valide based on its application in two complete official PNLD cycles: PNLD 2023 – Object 1 and PNLD 2024 – Object 1. Valide constitutes the designed artifact, while its architecture and governance mechanisms represent the proposed design contribution, and the retrospective analysis of institutional operation provides the empirical evaluation.

The evaluated objects correspond to digital collections composed of Student Books and Teacher Manuals, covering the curricular components of Portuguese Language, Mathematics, Science, History, and Geography.

In the PNLD 2023 – Object 1 cycle, 982 digital volumes from 111 collections submitted by 18 publishers were evaluated. In the PNLD 2024 – Object 1 cycle, 724 volumes organized into 90 collections, also originating from 18 publishers, were evaluated. Altogether, the study covers 1,706 digital volumes corresponding to the totality of works submitted to the technical evaluation stages of the respective cycles, without partial sampling.

5.1. Experimental Procedure and Unit of Analysis

Traditional manual evaluation was conducted by teams of human analysts following PNLD institutional procedures. In the PNLD 2023 cycle, 116 analysts participated, and in PNLD 2024, 91 analysts participated. In both cycles, the volumes underwent independent double-checking to increase consistency and reliability of results.

The same set of works was subsequently processed by Valide. Although the platform includes deterministic, AI-based, and hybrid validators, the analyzed dataset contains only findings generated by deterministic validators based on explicit PNLD rules and technical specifications. Therefore, the reported confirmation rates characterize only the deterministic validation layer.

The automatically generated findings are not treated as final decisions. Each occurrence is submitted to human validation performed by technology analysts responsible for confirming or discarding the findings according to the current regulatory criteria. In PNLD 2023, 49 technology analysts participated, and in PNLD 2024, 45 analysts participated.

The basic unit of analysis adopted in this study is the individual technical finding, defined as a potential nonconformity identified in a specific volume during technical evaluation. Because manual evaluation consolidates related nonconformities, its records are not directly comparable to the occurrence-level findings generated by Valide.

After human validation, the findings were classified as: (i) confirmed failure, when the finding effectively corresponds to a nonconformity; and (ii) non-confirmed failure, when the finding is discarded after human analysis.

5.2. Metrics and Analytical Strategy

The main metric used in this study is the *finding confirmation rate*, defined as:

$$ConfirmationRate = \frac{N_{confirmed}}{N_{confirmed} + N_{non-confirmed}} \quad (1)$$

This metric represents the proportion of automated findings confirmed after human review in a real regulatory context, indicating the degree of confirmation of the generated findings. It does not allow the estimation of false negatives, recall, or coverage of existing nonconformities.

The analysis focused on the longitudinal comparison between the two evaluated cycles, considering: (i) total volume of generated findings; (ii) proportion of confirmed and non-confirmed failures; (iii) distribution of findings by requirement category; and (iv) individual behavior of validators.

The investigation was guided by three analytical expectations: (i) strictly formal requirements would tend to present high agreement between automation and human validation; (ii) context-dependent requirements would present a greater need for human review; and (iii) confirmation rates would remain consistent despite changes in scale and in the number of validators.

6. Results

This section presents the results of the deterministic validators analyzed in the PNLD 2023 – Object 1 and PNLD 2024 – Object 1 cycles. The analysis considers the findings generated by deterministic validators, their subsequent human validation, and the relationship of these results with the consolidated outcomes of the traditional manual evaluation.

Table 3 summarizes the results of the two evaluated cycles. In PNLD 2023 – Object 1, the analyzed configuration included 43 deterministic validators and generated 48,967 technical findings, of which 47,307 were confirmed after human validation, resulting in a global confirmation rate of 96.6%. In PNLD 2024 – Object 1, the analyzed configuration included 52 deterministic validators, with 82,334 generated findings, 78,419 confirmed findings, and a global confirmation rate of 95.2%.

Tabela 3. Comparative summary of supervised deterministic validation results

Indicator	PNLD 2023 – Object 1	PNLD 2024 – Object 1
Evaluated volumes	982	724
Evaluated collections	111	90
Participating publishers	18	18
Active validators	43	52
Generated technical findings	48,967	82,334
Confirmed findings	47,307	78,419
Non-confirmed findings	1,660	3,915
Global confirmation rate	96.6%	95.2%
Global non-confirmation rate	3.4%	4.8%

Several validators associated with strictly formal and structural requirements achieved 100% confirmation rates in both cycles. Among them are validators related to the structural integrity of the digital package, absence of mandatory files, manifest conformity, presence of mandatory metadata, and correct declaration of the HTML5 standard, such as validators ae11, ae13c, ae20d, ae23a, ae24c, and ae24d. These results indicate high confirmability of findings generated for objective and deterministically formalizable requirements.

On the other hand, some validators presented lower confirmation rates, especially those associated with requirements more dependent on technical, editorial, or structural context. In the PNLD 2023 cycle, the validators `ac2a` (59.1%), related to the identification of single-item lists, and `ae19d` (27.3%), related to cover image dimensions, stood out. In the PNLD 2024 cycle, lower rates were also observed in validators related to the structural coherence of links and internal references, such as `ac7c` (82.4%) and `ae3` (87.5%).

Within the adaptive governance model adopted in Valide, these cases indicate the need for validator refinement. Validators with reduced confirmation rates return to cycles of technical review and adjustment of the implemented rules, involving analysis of false positives, refinement of detection conditions, and reevaluation of operational behavior in subsequent cycles. This process enables incremental evolution of validators without compromising traceability, auditability, or human supervision.

Across the two evaluated cycles, the global confirmation rates remained above 95% despite changes in the number of validators and generated findings. The variation of less than two percentage points provides preliminary evidence of consistency under the analyzed conditions, but additional cycles are needed to assess longitudinal stability.

Because manual evaluation consolidates related nonconformities while Valide records individual occurrences, their absolute numbers are not directly comparable. Findings generated by Valide include location markers and technical evidence for each occurrence.

This granularity supports targeted corrections and directs human review to specific occurrences, although the reduction in manual effort was not measured. Furthermore, validators that accumulate consistent operational evidence across multiple cycles may evolve toward higher automation regimes, while more context-dependent requirements remain subject to mandatory human validation.

Considering both cycles together, the results indicate confirmation rates above 95% in a large-scale regulatory context with high technical heterogeneity. The analysis suggests that strictly formal requirements tend to produce findings with higher confirmation rates, while context-dependent requirements continue to demand more intensive human supervision.

7. Discussion

This section discusses the contributions of Valide to the large-scale regulatory validation of digital educational resources and analyzes the methodological limitations and threats to the validity of the study.

7.1. Contributions to Digital Education, Accessibility, and Regulatory Governance

The results show that automated validators can support large-scale technical evaluation while maintaining human review and traceability. However, gains in time, productivity, cost, or human effort were not measured. The operational experience with Valide indicates that automated validation can be structured as an auditable and human-supervised process even in scenarios characterized by high editorial heterogeneity and large evaluation volumes.

Confirmation rates above 95% in both cycles indicate high confirmability of the findings generated by the analyzed deterministic validators. This result is consistent with the explicit separation between automated execution and final institutional decision-making, preserving human authority over regulatory interpretations, contextual exceptions, and the consolidation of results.

The model adopted by Valide assumes that the level of human validation may vary according to operational evidence. Strictly technical validators with stable behavior and high confirmation rates may gradually evolve toward higher automation regimes. In contrast, requirements dependent on contextual interpretation remain subject to mandatory human validation.

The similar confirmation rates observed in the PNLD 2023 and PNLD 2024 cycles provide preliminary evidence of consistency under the analyzed conditions. However, confirming the system's stability and robustness requires evaluation across additional cycles.

The occurrence-level granularity of Valide supports the localization of structural failures, accessibility problems, and regulatory inconsistencies and the production of traceable evidence, but prevents direct comparison with the aggregated records of the manual workflow. The findings also provide analysts and publishers with more precise information for review and corrections. However, the corresponding time and staffing requirements were not quantified.

Valide should not be interpreted as a replacement for specialized tools such as EPUBCheck or ACE, nor does this study demonstrate superior detection accuracy. While these tools focus on specific structural or accessibility standards, Valide integrates validation mechanisms with regulatory versioning, rule-level evidence, human review, QA, and institutional governance. Its main distinction therefore lies in the orchestration of validation within an auditable regulatory workflow.

From the perspective of AI in Education, Valide illustrates how an augmented intelligence infrastructure can combine deterministic, AI-based, and hybrid validation mechanisms with human review and institutional governance to support large-scale digital educational ecosystems. The empirical evidence reported in this study concerns the deterministic validation layer, while the specific evaluation of AI-based and hybrid components remains future work. Unlike approaches focused exclusively on full automation, Valide adopts human-centered AI principles, in which these mechanisms support human institutional judgment. This approach aligns with contemporary principles of trustworthy AI and algorithmic governance, including transparency, auditability, human oversight, and institutional accountability [Yang et al. 2021, Wirtz et al. 2019].

By identifying structural, technical, and accessibility problems, Valide may indirectly support the quality of the digital educational resources used by students and teachers in the PNLD. The systematic identification of structural failures, navigation problems, technical inconsistencies, and accessibility barriers can support usability, readability, and equitable access to digital educational resources. In particular, automated verification of accessibility requirements supports the early detection of problems that may compromise access to pedagogical content for students with disabilities. Thus, the system operates not only as a regulatory compliance mechanism, but also as an infrastructure

supporting quality, inclusion, and accessibility in large-scale digital education.

In the context of educational public policies, systems such as Valide become particularly relevant due to the continuous growth in the volume, technical diversity, and regulatory complexity of digital educational resources. The production of traceable evidence and the systematic identification of failures can support national digital education programs that require accessibility, technical compliance, and institutional transparency.

7.2. Threats to Validity

This study presents limitations that should be considered when interpreting the results. Regarding internal validity, no controlled experiment compared manual and Valide-supported workflows under equivalent conditions. Therefore, differences in execution time, productivity, cost, or human effort were not measured, limiting causal claims about the system's impact. In addition, the confirmation rate does not account for undetected nonconformities. Future work will estimate recall and coverage through independently annotated samples or mappings between manual and automated findings at the work-and-requirement level.

Regarding external validity, the results are associated with the institutional context of PNLD Digital, which has specific regulatory characteristics. The empirical dataset is restricted to deterministic validators and therefore does not characterize the performance of the AI-based or hybrid validators implemented in Valide.

8. Conclusion and Future Work

This article presented Valide, an augmented intelligence infrastructure developed to support the large-scale regulatory validation of digital textbooks through the integration of modular automation, institutional governance, and supervised human validation.

The empirical analysis of two official PNLD Digital cycles indicated confirmation rates above 95%, showing that deterministic validators can support large-scale technical inspection within a traceable and human-supervised workflow.

Although developed in the context of the PNLD, the proposed principles, such as auditable validators, production of verifiable evidence, and separation between automated detection and institutional decision-making, have potential applicability in other scenarios involving regulatory evaluation of digital artifacts. The work also contributes to the debate on trustworthy AI and *human-centered AI* in Education, demonstrating the potential of augmented intelligence infrastructures to support complex digital educational ecosystems.

As future work, controlled experimental studies comparing manual and automated workflows are intended, as well as expansion of the analysis to hybrid and AI-based validators for requirements that are more dependent on contextual interpretation.

Referências

- Abduganiev, S. G. (2017). Towards automated web accessibility evaluation: a comparative study. *International Journal of Information Technology and Computer Science*, 9(9):18–44.
- Alsaedi, A. (2020). Comparing web accessibility evaluation tools and evaluating the accessibility of webpages: Proposed frameworks. *Information*, 11(1):40. Open access (MDPI).

- DAISY Consortium (2024a). ACE by DAISY. <https://daisy.org/activities/software/ace/>. Accessibility checking tool for EPUB.
- DAISY Consortium (2024b). EPUBCheck. <https://github.com/w3c/epubcheck>. Open-source validation tool.
- Gay, G. and Li, C. Q. (2010). Achecker: open, interactive, customizable, web accessibility checking. In *Proceedings of the 2010 International cross disciplinary conference on web accessibility (W4A)*, pages 1–2.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly*, 28(1):6.
- Larson, L. C. (2010). Digital readers: The next chapter in e-book reading and response. *The reading teacher*, 64(1):15–22.
- Marques, L. B., Sobrinho, Á. A. d. C. C., da Silva, A. P., and Rodrigues, N. C. I. (2023). Esforços para a conformidade com critérios de acessibilidade no programa nacional do livro e do material didático. *Cadernos do FNDE*, 4(09):e0981–e0981.
- Ozbay, I. and Ugurelli, Y. O. (2023). Changing children’s literature in the digital age: Digital books. *International Journal of Education and Literacy Studies*, 11(1):68–85.
- Padure, M. and Pribeanu, C. (2020). Comparing six free accessibility evaluation tools. *Informatica Economica*, 24(1):15–25.
- Paiva, R., Xisto, J., Sobrinho, Á., Silva, A., Sarmiento, F., Recch, F., Tenório, S., Carvalho, A., Bittencourt, I., and Isotani, S. (2025). Expanding the resilience of the brazilian education system by supporting the evaluation of digital textbooks. *Humanities and Social Sciences Communications*, 12(1):1–9.
- Sobrinho, Á., Ibert Bittencourt, I., Carvalho Melo da Silveira, A., Pedro da Silva, A., Dermeval, D., Brandão Marques, L., Cezar Ianzer Rodrigues, N., Carolina Silva e Souza, A., Ferreira, R., and Isotani, S. (2023). Towards digital transformation of the validation and triage process of textbooks in the brazilian educational policy. *Sustainability*, 15(7):5861.
- Tenório, S., Sobrinho, Á., Rosa, D., Paiva, R., Marques, L., Dermeval, D., da Silva, A. P., Cordeiro, T., Isotani, S., and Bittencourt, I. I. (2025). Artificial intelligence and public education policy: Improving textbook evaluation in brazil. In *International Conference on Artificial Intelligence in Education*, pages 208–220. Springer.
- UNESCO, P. (2021). *Reimagining our futures together: A new social contract for education*. Educational and Cultural Organization of the United Nations Paris, France.
- Vigo, M., Brown, J., and Conway, V. (2013). Benchmarking web accessibility evaluation tools: Measuring the harm of sole reliance on automated tests. *Journal of Web Engineering*, 12(3–4):207–226. Versões abertas disponíveis via repositórios institucionais.
- Wirtz, B. W., Weyerer, J. C., and Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, 42(7):596–615.
- Xu, Y. and Gao, B. (2025). Trust and ai weight: human-ai collaboration in organizational management decision-making. *Frontiers in Organizational Psychology*, 3.

Yang, S. J., Ogata, H., Matsui, T., and Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2:100008.

Zambon, L. B. and Terrazzan, E. A. (2013). Políticas de material didático no brasil: organização dos processos de escolha de livros didáticos em escolas públicas de educação básica. *Revista Brasileira de Estudos Pedagógicos*, 94(237):585–602.