

Do papel ao dado: escanear, transcrever, organizar

A escola básica brasileira aprende no papel, e o celular saiu da aula. O hackeamento da ensinagem começa transformando a prova, a lição e a redação manuscritas em dado legível, sem perder o que só o professor vê. Cinco etapas, da correção por questão à proteção da imagem, e o que fica com o humano em cada uma.

1

Corrigir por questão

Ler a mesma questão de todos os alunos, com rubrica viva e respostas agrupadas, em vez de prova por prova.

2

Transcrever o manuscrito

Modelos multimodais já leem a letra do aluno; o gargalo é a transcrição, e os desenhos são lidos pior.

3

O português manuscrito

Redações escolares com rasuras e trechos ilegíveis são o caso difícil, e já há base pública anotada para medi-lo.

4

O que fica com o humano

O professor confere toda transcrição antes de qualquer devolutiva, e não só o ilegível, o rasurado e o desenho.

5

A imagem é dado pessoal

Prazo para apagar o escaneamento, nenhuma identificação por caligrafia, dados no Brasil e anonimizados.

O papel é o sensor que a escola já tem

A literatura mede cliques; a sala de aula brasileira produz folhas.

O SENSOR QUE FALTA

A learning analytics nasceu medindo rastros de ambientes virtuais no ensino superior. Na educação básica brasileira, a aprendizagem acontece em cadernos, provas e folhas de redação, e a Lei 15.100/2025 tirou o celular da sala. O papel escaneado leva à análise **o que o aluno de fato produziu**.

MAIS RICO QUE O CLIQUE

Uma folha de prova guarda a resposta, o cálculo, a rasura, a tentativa abandonada; um registro de acesso guarda o horário. Para quem busca **a regra do erro** (material 3, *O erro tem regra*), o caminho do aluno no papel vale mais que qualquer métrica de uso.

O CUSTO DO SENSOR

O papel não chega pronto à análise: precisa ser escaneado, identificado, transcrito e conferido, e cada etapa pode errar. O erro de transcrição contamina tudo o que vem depois. Por isso a **qualidade do dado** vem antes da leitura por critérios: é o "instrumentar antes de medir" de *Engenharia da aprendizagem*, material 3.

No IA para escolas. É a demonstração proposta ao LAK27 (Learning Analytics and Knowledge, Recife, março de 2027): "do manuscrito digitalizado à leitura por critérios". Um PDF por turma, identificação pelo cabeçalho ou pela ordem da chamada, a foto da lição pelo celular e a transcrição conferida pelo professor.

Corrigir por questão

A mesma questão, de todos, de uma vez.

O QUE É

O Gradescope (Singh e col., 2017) inverteu a ordem da correção: em vez de prova por prova, **questão por questão**, com uma rubrica que pode mudar durante o trabalho e se reaplica a quem já foi corrigido. Ler a mesma questão da turma inteira revela padrões que a pilha de provas esconde.

O AGRUPAMENTO

Respostas iguais ou parecidas são agrupadas, e no Gradescope uma decisão vale para o grupo inteiro. Ganha-se tempo e coerência; perde-se a chance de notar **a diferença dentro do grupo**, aquela resposta que parece igual às outras e não é.

A DIFERENÇA

Aqui o agrupamento só orienta a atenção. A IA corrige os itens objetivos, mas o professor valida **cada resultado**, com registro, antes do lançamento, nunca em bloco. É a regra da correção validada no *Fundamento 4, Tipos de correção com IA*.

No IA para escolas. O supergabarito é gerado a partir do gabarito simples, e as respostas idênticas aparecem agrupadas para revisão. A IA pode destacar casos sinalizados e notas baixas para orientar o olhar; a decisão continua sendo tomada resultado a resultado.

2 Transcrever o manuscrito

Ler a letra do aluno é a etapa mais frágil.

A EVIDÊNCIA

Num estudo com 3.945 registros de testes curtos de cálculo manuscritos, cerca de **88%** das transcrições feitas por modelos multimodais foram aceitáveis (Yu e col., 2026). É muito, e não é tudo: a cada cem registros, uns doze precisam de reparo antes de qualquer leitura.

O GARGALO

Numa prova manuscrita de termodinâmica com 252 alunos, o ponto fraco da correção assistida foi **a transcrição**, e não o julgamento; diagramas desenhados à mão foram avaliados com menos confiabilidade que as deduções matemáticas (Kortemeyer, Nöhl e Onishchuk, 2024).

A RÉGUA HUMANA

Os humanos também divergem: no estudo de cálculo, dois revisores humanos concordaram exatamente em cerca de 45% das questões (Yu e col., 2026). A conferência humana não garante acerto; garante que **alguém responde** pela leitura.

No IA para escolas. A transcrição por OCR (reconhecimento óptico de caracteres) é conferida pelo professor antes de qualquer leitura por critérios. Regra prática da escola: fórmulas, esquemas e desenhos se leem na imagem original, porque é ali que a máquina mais erra.

O português manuscrito

Rasura, letra cursiva e trecho ilegível.

A BASE

O BRESSAY (Neto e col., 2024), apresentado na ICDAR (International Conference on Document Analysis and Recognition), reúne **1.000 páginas** de redações manuscritas em português, cada uma de um autor, transcritas por linha, palavra, parágrafo e página. É uma régua pública para a letra escolar brasileira.

O QUE SE ANOTA

A base marca o que a escola conhece bem: **texto ilegível**, inserções acima ou abaixo da linha, palavras riscadas que ainda se leem e as que já não se leem. Esses casos não são ruído; são onde a transcrição automática mais tropeça e onde o professor mais precisa olhar.

O VIÉS DA LETRA

A letra difícil pesa na nota humana: é o viés da letra do *Fundamento 5*, *Viés humano* e *viés algorítmico*. A transcrição conferida pode **reduzi-lo**, porque o professor lê o texto, e não só a caligrafia; mas a dúvida de leitura precisa ficar registrada, e não resolvida em silêncio.

No IA para escolas. Na leitura por critérios das redações, o professor confere a transcrição contra a imagem e lê cada redação por inteiro antes de atribuir a nota; a IA não atribui nota. Trecho ilegível se marca como ilegível: não se adivinha o que o aluno quis escrever.

4 O que fica com o humano

Toda transcrição, não só a duvidosa.

O QUE É

O professor confere **toda** transcrição antes de qualquer devolutiva, e não apenas o trecho ilegível, o rasurado ou o desenho. Erros de transcrição podem parecer plausíveis: uma vírgula trocada muda a pontuação, um sinal trocado muda a conta, e nada avisa do problema.

POR QUE TUDO

Conferir só o que a máquina marca como dúvida é confiar na máquina para saber onde ela erra. O hábito 4 da *Fricção cognitiva ótima e aprendizagem* diz: quanto maior o custo do erro, **maior a fricção**. Um erro de leitura vira nota injusta e devolutiva equivocada.

O REGISTRO

Vale registrar cada correção de transcrição. Os registros mostram onde a leitura automática falha (que letra, que tipo de questão) e viram evidência para decidir onde ela é confiável. É o que *Engenharia da aprendizagem* (material 3) chama de **instrumentar** antes de medir.

No IA para escolas. A ordem é fixa: escanear, identificar, transcrever, conferir e só então ler por critérios. Nos itens objetivos, cada resultado é validado um a um; nas redações, a nota é do professor, e o aluno só vê o que ele liberou.

5 A imagem é dado pessoal

Escanear também é tratar dados de menores.

O QUE É

A folha escaneada traz nome, letra, rasuras e às vezes desenhos e anotações pessoais. É dado pessoal de criança ou adolescente, protegido pela LGPD (Lei Geral de Proteção de Dados), art. 14, que exige tratamento no **melhor interesse** do titular.

COMO SE FAZ

Finalidade escrita, prazo para apagar o escaneamento, acesso restrito a quem corrige. **Nenhuma identificação por caligrafia:** a letra serve para ler o texto, nunca para reconhecer o autor. Quem é o aluno se sabe pelo cabeçalho ou pela ordem da chamada.

A RÉGUA LEGAL

Valem também o ECA Digital (Estatuto Digital da Criança e do Adolescente, Lei 15.211/2025, em vigor desde março de 2026) e as normas educacionais sobre IA do CNE (Conselho Nacional de Educação) e do MEC (Ministério da Educação), comparadas no Guia *Divergências CNE/MEC*. O Guia (Parte II) põe dados sensíveis no **alto risco**.

No IA para escolas. Os dados ficam hospedados no Brasil e anonimizados por identificadores (IDs), com exportação em JSON e exclusão em 15 dias úteis. Nenhum dado é usado para publicidade, fins comerciais ou treinamento de modelos.

Seis condições para o papel virar dado

- ✓ O escaneamento está nítido, completo e na ordem certa, com uma turma por PDF.
- ✓ O professor conferiu toda a transcrição contra a imagem, e não só os trechos marcados como dúvida.
- ✓ Trechos ilegíveis estão marcados como ilegíveis, e não adivinhados.
- ✓ Cada folha está ligada ao aluno certo pelo cabeçalho ou pela ordem da chamada, nunca pela letra.
- ✓ Fórmulas, esquemas e desenhos foram lidos na imagem original.
- ✓ Há prazo escrito para apagar as imagens, e ninguém fora da correção tem acesso a elas.

Créditos e referências

Segundo material da série *Hackeamento da ensinagem*. Depois do conceito (material 1), trata do sensor: como a produção em papel vira dado confiável. Dialoga com o viés da letra (*Fundamento 5*), a correção validada (*Fundamento 4*), o fluxo de correção da Apresentação 28p e o "instrumentar antes de medir" de *Engenharia da aprendizagem*. A série *Learning analytics* lê o mesmo dado como item (material 4) e como redação (material 5); o material 3 desta série, *O erro tem regra*, começa onde este termina.

SINGH, A. et al. Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten work. In: L@S '17. New York: ACM, 2017. p. 81–88.

YU, Z. et al. Evaluating AI grading on real-world handwritten college mathematics: a large-scale study toward a benchmark. arXiv:2603.00895, 2026.

KORTEMAYER, G.; NÖHL, J.; ONISHCHUK, D. Grading assistance for a handwritten thermodynamics exam using artificial intelligence: an exploratory study. Physical Review Physics Education Research, v. 20, n. 2, 020144, 2024.

NETO, A. F. S. et al. BRESSAY: a Brazilian Portuguese dataset for offline handwritten text recognition. In: BARNEY SMITH, E. H.; LIWICKI, M.; PENG, L. (Ed.). Document analysis and recognition: ICDAR 2024, Part II. Cham: Springer, 2024. p. 315–333. (Lecture Notes in Computer Science, v. 14805).

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018.

BRASIL. Lei nº 15.211, de 17 de setembro de 2025. Dispõe sobre a proteção de crianças e adolescentes em ambientes digitais (Estatuto Digital da Criança e do Adolescente). Diário Oficial da União: seção 1, edição extra, Brasília, DF, 17 set. 2025.

BRASIL. Lei nº 15.100, de 13 de janeiro de 2025. Dispõe sobre a utilização, por estudantes, de aparelhos eletrônicos portáteis pessoais nos estabelecimentos públicos e privados de ensino da educação básica. Diário Oficial da União: seção 1, Brasília, DF, 14 jan. 2025.