<table>
<tr><td>Título do plano</td><td colspan="3">Definições do Dataset</td><td>Sequência:
8</td></tr>
<tr><td>Etapa/Ano</td><td colspan="4">Ensino Médio: ☐ 1º Ano | ☐ 2º Ano | ☑ 3º Ano</td></tr>
<tr><td>Nível de Maturidade
[link]</td><td>Escola</td><td>☐ Emergente
☐ Básico
☑ Intermediário
☐ Avançado</td><td>Docente</td><td>☐ Básico
☑ Intermediário
☐ Avançado</td></tr>
<tr><td>Competências
[link]</td><td colspan="4">☐ C1 ☑ C2 ☑ C3 ☑ C4 ☑ C5</td></tr>
<tr><td>Objetos de Conhecimento</td><td colspan="4">• Documentação de dados em projetos de IA;
• Definição formal de um dataset;
• Origem dos dados: fonte, instrumento, contexto e responsáveis pela coleta;
• Metadados: data, local, público respondente, versão do formulário, versão da base e responsável pela curadoria;
• Limitações do dataset: amostragem, representatividade, faltantes, duplicatas, categorias, vieses e restrições de interpretação;
• Uso permitido e uso não recomendado dos dados;
• Privacidade, confidencialidade e responsabilidade no uso de dados escolares;
• Versionamento da base: base original, base limpa, base documentada e base pronta para análise/modelagem;
• Relação entre documentação do dado, transparência e explicabilidade em IA;
• Diferença entre documentar dados e documentar modelos;
• Fichas de documentação de dados, inspiradas na ideia de datasheets for datasets;
• Registro de decisões técnicas e pedagógicas sobre o dataset;
• Condições para uso dos dados em análise, visualização e treinamento de modelos de IA.</td></tr>
<tr><td>Recursos Educacionais</td><td colspan="4">• Questionário padrão do projeto Vozes da Escola ou projeto regional;
• Base sintética de respostas já estruturada, coletada e limpa nos planos anteriores;
• Planilha eletrônica, como Google Planilhas, Excel ou LibreOffice Calc;
• Modelo de ficha de documentação do dataset;
• Modelo de dicionário de dados;
• Quadro, projetor ou cartolina;
• Documento colaborativo para preenchimento coletivo;
• Cópia da base original e da base tratada;
• Registro das decisões de limpeza produzido no plano 307;
• Lista oficial das iniciativas I01 a I12.</td></tr>
<tr><td>Componentes Curriculares Relacionados</td><td>☐ Arte
☐ Língua Port.
☐ Ed. Física</td><td>☐ Geografia
☐ História
☐ Matemática</td><td colspan="2">☐ Língua Inglesa
☐ Ciências
☑ Computação</td></tr>
<tr><td>Palavras-chave</td><td colspan="4">Dataset, documentação de dados, metadados, origem dos dados, limitações, uso permitido, versionamento, transparência, curadoria, IA responsável.</td></tr>
</table>

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Perguntas Importantes	● O que é necessário registrar para que outras pessoas compreendam um dataset? ● De onde vieram os dados? ● Quem respondeu ao questionário? ● Quem ficou de fora da coleta? ● Quais limitações a base apresenta? ● O que pode e o que não pode ser concluído a partir desses dados? ● Que usos são permitidos para esta base? ● Que usos seriam inadequados ou injustos? ● Por que a base original deve ser preservada? ● Qual é a diferença entre a versão bruta, limpa e documentada do dataset? ● Como a documentação ajuda a evitar interpretações erradas? ● Como a documentação do dataset melhora a transparência em projetos de IA? ● O que pode acontecer quando um sistema de IA usa dados sem documentação?
Encontros	1 **Total de Horas-aula** 2
Objetivos	● Compreender que a definição de um dataset não se limita à planilha de dados; ● Identificar os principais elementos que devem ser documentados em uma base de dados; ● Registrar origem, contexto, responsáveis, instrumento de coleta e versão da base; ● Reconhecer limitações do dataset relacionadas à coleta, amostragem, limpeza e representatividade; ● Definir usos permitidos, usos não recomendados e cuidados éticos no uso dos dados; ● Compreender a importância do versionamento para rastrear alterações na base; ● Relacionar a documentação do dado com transparência, responsabilidade e confiabilidade em IA; ● Produzir uma ficha de documentação do dataset do projeto Vozes da Escola ou projeto regional.
Habilidades Relacionadas	**BNCC Computação:** (EM13CO12) Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados. (EM13CO13) Analisar e utilizar as diferentes formas de representação e consulta a dados em formato digital para pesquisas científicas. interpretar a base. (EM13CO14) Avaliar a confiabilidade das informações encontradas em meio digital, investigando seus modos de construção e considerando autoria, estrutura e propósito da mensagem. (EM13CO18) Planejar e gerenciar projetos integrados às áreas de conhecimento de forma colaborativa, solucionando problemas, usando diversos artefatos computacionais.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

	Referencial Curricular IA no EM: (EM02IA06) Compreender o ciclo de projeto envolvido com o desenvolvimento de sistemas de IA. (EM03IA03) Saber treinar uma IA por meio da técnica de Machine Learning. (EM03IA05) Compreender que a IA resulta da qualidade e da diversidade das bases de dados com os quais interage. **Habilidades articuladoras da progressão anterior:** (EF69IA08) Aplicar o ciclo de vida de um sistema de IA, criando uma solução que considere princípios éticos desde a concepção até a implementação, garantindo transparência, equidade e responsabilidade em todas as etapas. A documentação do dataset é uma prática concreta de transparência, equidade e responsabilidade no ciclo de vida de IA. (EF69IA20) Utilizar dados com responsabilidade, garantindo a privacidade e demonstrando compromisso pessoal e social em sociedades impulsionadas pela IA. A aula mobiliza essa habilidade ao discutir uso permitido, privacidade, limitações e riscos de reutilização inadequada dos dados escolares.
Práticas Pedagógicas Inovadoras	☑ Aula Enriquecida com Tecnologia ☐ Ensino Híbrido: Sala de aula invertida ☐ Ensino Híbrido: rotação por estação ☐ Ensino Híbrido: rotação individual ☑ Aulas Mão na massa ☑ Aprendizagem baseada em projetos ☐ IA Desconectada

Sequência Didática

1. Introdução ao tema

A aula pode iniciar com a pergunta:

"Se outra pessoa recebesse nossa planilha amanhã, ela entenderia de onde vieram os dados, o que eles significam e como poderiam ser usados?"

O professor retoma brevemente os planos anteriores:

- no Plano 304, os estudantes compreenderam que um dataset precisa de estrutura;
- no Plano 305, analisaram o questionário padrão como instrumento de coleta;
- no Plano 306, discutiram coleta, amostragem e representatividade;
- no Plano 307, trabalharam a limpeza dos dados e o registro das decisões tomadas.

Agora, a turma deve compreender que, mesmo com dados estruturados, coletados e tratados, ainda é necessário **documentar o dataset**.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Sugestões de apontamentos:

- Uma planilha não fala sozinha.
- O dataset precisa contar sua história: de onde veio, como foi coletado, como foi tratado e para que pode ser usado.
- Documentar dados é uma prática de transparência e responsabilidade.
- Em IA, dados mal documentados podem gerar análises equivocadas, modelos pouco confiáveis ou usos inadequados.

Perguntas para iniciar a discussão:

- Quem produziu esses dados?
- Quando eles foram coletados?
- Quem respondeu ao questionário?
- Quem ficou de fora?
- Que decisões de limpeza foram tomadas?
- A base pode ser usada para treinar uma IA?
- Que cuidados seriam necessários antes de usar essa base em um modelo?
- O que uma pessoa externa precisaria saber para interpretar corretamente os dados?

2. O que significa definir um dataset?

O professor apresenta a ideia de que um dataset não é apenas uma tabela com respostas. Ele é um conjunto de dados com finalidade, estrutura, origem, contexto, limitações e condições de uso.

Sugestões de explicações:

- Definir o dataset é explicar o que ele é.
- Documentar o dataset é registrar informações que ajudam outras pessoas a compreenderem e usarem a base corretamente.
- Uma base pode estar limpa, mas ainda assim ser mal utilizada se não estiver documentada.
- A documentação ajuda a evitar conclusões que os dados não sustentam.

Exemplo simples:

Uma planilha com respostas do projeto Vozes da Escola não deve ser apresentada apenas como “dados da escola”. Ela precisa ser descrita como:

- dados coletados por meio do questionário padrão;
- respondidos por determinados grupos da comunidade escolar;
- coletados em uma janela específica;
- tratados conforme critérios definidos;
- destinados a apoiar a priorização coletiva de iniciativas;

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

- não destinados a avaliar individualmente estudantes, professores, famílias ou turmas.

Perguntas para discussão:

- O que muda quando uma planilha passa a ter uma ficha de documentação?
- Como saber se um dado pode ser usado para uma finalidade específica?
- O que a documentação revela que a planilha sozinha não mostra?
- Que riscos aparecem quando olhamos apenas para os números?

3. Origem dos dados: de onde vieram?

Esta etapa retoma o Plano 306, que discutiu coleta, amostragem, janela de coleta e contexto dos dados. O professor apresenta a primeira parte da ficha de documentação: **origem dos dados**.

Sugestão de campos:

Campo	O que registrar
Nome do dataset	Nome claro da base
Projeto	Nome do projeto ou atividade
Finalidade da coleta	Objetivo da coleta
Instrumento de coleta	Questionário utilizado
Responsáveis pela coleta	Turma, professor, escola ou equipe
Período de coleta	Data inicial e data final
Local/contexto	Escola, turma, reunião, evento ou aplicação online
Público-alvo	Quem se desejava ouvir
Forma de aplicação	Online, presencial, mediada, individual ou coletiva
Versão do questionário	Exemplo: formulário_v1.0
Observações da coleta	Dificuldades, baixa adesão, grupos menos representados

Sugestões de explicações:

- A origem dos dados ajuda a entender o contexto da base.
- Duas bases com o mesmo questionário podem ter resultados diferentes se a coleta ocorreu em públicos, momentos ou condições diferentes.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

- Em IA, saber a origem dos dados é importante para avaliar se eles fazem sentido para o problema que queremos analisar.

Perguntas para discussão:

- Quem respondeu ao questionário?
- Quem talvez não tenha conseguido responder?
- A coleta foi feita em um momento adequado?
- A forma de aplicação pode ter influenciado as respostas?
- Que informação importante sobre a coleta não aparece diretamente na planilha?

4. Limitações do dataset: o que os dados não mostram?

Esta é uma das etapas centrais da aula.

O professor explica que documentar limitações não diminui o valor da base. Pelo contrário, torna a análise mais honesta e confiável.

Sugestões de explicações:

- Todo dataset tem limitações.
- Uma limitação não significa que a base é inútil.
- A limitação indica até onde podemos ir com a interpretação.
- Em IA, reconhecer limitações evita usos indevidos e conclusões exageradas.

Exemplos de limitações possíveis:

Tipo de limitação	Exemplo
Amostragem	A maior parte das respostas veio de estudantes
Representatividade	Poucos familiares responderam ao questionário
Janela de coleta	A coleta ocorreu em poucos dias
Acesso	Pessoas sem internet ou celular podem ter participado menos
Faltantes	Alguns campos opcionais ficaram sem resposta
Duplicatas	Algumas respostas foram removidas da base tratada
Categorias	Algumas respostas precisaram ser padronizadas
Interpretação	A base indica percepções, não comprova causa e efeito
Uso em IA	A base não deve ser usada para decisões automáticas sobre pessoas

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Perguntas para discussão:

- Que grupos estão mais representados na base?
- Que grupos aparecem pouco?
- O que os dados permitem afirmar?
- O que os dados não permitem afirmar?
- Seria correto dizer que a iniciativa mais votada é “a melhor” para todos?
- Seria justo usar esses dados para avaliar o desempenho de uma turma ou professor?
- O que uma IA poderia interpretar errado se essas limitações não fossem registradas?

Mensagem importante:

> Documentar limitações é uma prática de transparência. A limitação não invalida a base; ela orienta seu uso responsável.

5. Uso permitido e uso não recomendado

Nesta etapa, a turma discute a diferença entre usar dados para apoiar decisões coletivas e usar dados para tomar decisões automáticas, individualizantes ou injustas.

Sugestões de explicações:

- Nem todo uso tecnicamente possível é pedagogicamente adequado.
- A base foi construída para apoiar uma leitura coletiva da comunidade escolar.
- Ela não deve ser usada para vigiar, classificar ou punir indivíduos.
- O uso permitido precisa estar coerente com a finalidade original da coleta.

Quadro de apoio:

Uso permitido	Uso não recomendado
Identificar prioridades gerais da comunidade escolar	Avaliar individualmente estudantes
Produzir gráficos e sínteses coletivas	Comparar professores ou turmas como ranking
Apoiar planejamento de iniciativas escolares	Tomar decisões automáticas sem discussão humana
Discutir percepções de diferentes grupos	Expor respostas individuais
Apoiar projetos de aprendizagem com IA	Treinar modelo para identificar ou rotular pessoas

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Analisar tendências de participação	Usar os dados fora do contexto escolar sem autorização

Perguntas para discussão:

- Para que finalidade esses dados foram coletados?
- Que usos respeitam essa finalidade?
- Que usos poderiam prejudicar pessoas ou grupos?
- O que deve ser feito antes de compartilhar a base com outra turma ou escola?
- Que informações precisam ser removidas ou protegidas?

Mensagem importante:

> O uso dos dados precisa respeitar a finalidade da coleta, a privacidade dos participantes e a interpretação coletiva dos resultados.

6. Versionamento: por que controlar versões?

Esta etapa retoma o Plano 307, que trabalhou a preservação da base original, a criação da base tratada e o registro das decisões de limpeza.

O professor apresenta a ideia de versionamento.

Sugestões de explicações:

- Versionar é registrar as diferentes etapas da base.
- A base original deve ser preservada.
- A base tratada deve indicar o que foi alterado.
- A base documentada deve deixar claro que versão está sendo usada.
- Sem versionamento, não conseguimos saber de onde veio um resultado.

Modelo simples de versões:

Versão	Nome sugerido	Descrição
v0.1	base_original	Base exportada diretamente do formulário
v0.2	base_tratada	Base com faltantes, duplicatas e categorias tratadas
v1.0	base_documentada	Base tratada com ficha de documentação
v1.1	base_analise	Base preparada para gráficos e análises exploratórias

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

v2.0	base_modelagem	Base preparada para testes de modelo, se autorizado

Perguntas para discussão:

- Por que não devemos alterar diretamente a base original?
- Como saber qual versão foi usada para fazer um gráfico?
- O que acontece se dois grupos trabalham com versões diferentes?
- Como o versionamento ajuda a corrigir erros?
- Por que isso importa em IA?

Mensagem importante:

> Em projetos de dados e IA, resultados dependem da versão da base usada.

7. Relação com IA

Esta etapa explicita a relação entre documentação de dados e IA.

Sugestões de explicações:

- Modelos de IA aprendem padrões a partir de dados.
- Se os dados não têm origem clara, não sabemos exatamente o que o modelo aprendeu.
- Se as limitações não são documentadas, o modelo pode ser usado em contextos para os quais não foi preparado.
- Se o uso permitido não está definido, os dados podem ser reutilizados de forma inadequada.
- Se a versão da base não é registrada, não conseguimos reproduzir ou revisar os resultados.

Analogia:

> A documentação do dataset é como uma ficha técnica da base de dados.

Ela informa:

- quem produziu os dados;
- como foram coletados;
- o que representam;
- quais problemas ou limitações existem;
- quais usos são adequados;
- quais usos devem ser evitados;
- qual versão foi usada.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Perguntas para discussão:

- Um modelo treinado com essa base poderia ser usado em outra escola?
- O que precisaria ser documentado antes disso?
- A base representa bem todos os grupos da comunidade?
- Que risco existe se uma IA sugerir prioridades sem mostrar as limitações dos dados?
- Como a documentação ajuda professores, estudantes e gestores a manterem controle humano sobre o processo?

8. Atividade prática: criando a ficha de documentação do dataset

A turma será organizada em grupos. Cada grupo receberá:

- a base sintética do projeto Vozes da Escola ou projeto regional;
- o questionário padrão;
- o registro de limpeza do Plano 307;
- um modelo de ficha de documentação.

Cada grupo deverá preencher uma ficha com os seguintes blocos:

Bloco 1 — Identificação do dataset

Campo	Preenchimento
Nome do dataset	
Projeto	
Escola/turma	
Responsável pelo preenchimento	
Data da documentação	
Versão do dataset	

Bloco 2 — Origem e coleta

Campo	Preenchimento
Instrumento usado	
Período de coleta	
Público-alvo	
Grupos participantes	

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Forma de aplicação	
Número de respostas brutas	
Número de respostas após limpeza	
Observações sobre a coleta	

Bloco 3 — Estrutura da base

Campo	Preenchimento
O que cada linha representa?	
O que cada coluna representa?	
Quais campos são de perfil?	
Quais campos são prioridades?	
Quais campos são critérios de avaliação?	
Quais campos são texto livre?	
Quais campos são apoio/metadados?	

Bloco 4 — Limitações

Tipo de limitação	Observação
Amostragem	
Representatividade	
Faltantes	
Duplicatas	
Categorias padronizadas	
Campos opcionais	
Privacidade	
Interpretação dos resultados	

Bloco 5 — Uso permitido

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Tipo de uso	Permitido?	Observações
Gráficos e análise exploratória	Sim	
Planejamento coletivo de iniciativas	Sim	
Apresentação para a comunidade escolar	Sim, com cuidado	
Treinamento de modelo de IA	Somente com autorização e mediação	
Identificação individual de participantes	Não	
Ranking de turmas, professores ou grupos	Não	
Compartilhamento público da base completa	Não recomendado	

Bloco 6 — Versionamento

Versão	Nome da base	Descrição	Responsável	Data
v0.1	base_original	Exportada do formulário		
v0.2	base_tratada	Tratamento de faltantes, duplicatas e categorias		
v1.0	base_documentada	Base tratada com documentação		

9. Socialização: comparando fichas de documentação

Cada grupo apresenta uma parte da ficha preenchida.

Sugestões de perguntas para conduzir a socialização:

- Todos os grupos descreveram a origem dos dados da mesma forma?
- Quais limitações apareceram com mais frequência?
- Houve divergência sobre usos permitidos?
- Algum grupo identificou um risco de uso inadequado que os demais não perceberam?
- A ficha ajuda a entender melhor a base?
- O que ainda precisaria ser documentado antes de usar esses dados em IA?

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

O professor deve destacar que documentar dados é uma forma de tornar o projeto mais transparente, auditável e responsável.

10. Fechamento

Ao final, a turma constrói coletivamente um checklist mínimo de documentação do dataset.

Sugestão:

Item	Pergunta de verificação
Nome do dataset	A base tem um nome claro?
Finalidade	Está claro por que os dados foram coletados?
Origem	Está registrado de onde vieram os dados?
Instrumento	O questionário usado está identificado?
Público	Está claro quem respondeu?
Ausências	Está claro quem pode ter ficado de fora?
Limitações	As principais limitações foram registradas?
Limpeza	As decisões de tratamento foram documentadas?
Uso permitido	Está claro para que a base pode ser usada?
Uso não recomendado	Está claro o que não deve ser feito com os dados?
Versão	A versão da base está identificada?
Privacidade	Os riscos de exposição individual foram considerados?
IA	Está claro se a base pode ou não ser usada em modelos de IA?

Sugestões para encerramento:

- Documentar um dataset é explicar sua origem, seu contexto, seus limites e seus usos possíveis.
- A documentação não é burocracia: ela protege a interpretação dos dados.
- Em IA, uma base sem documentação pode gerar resultados pouco confiáveis ou usos inadequados.
- Antes de perguntar o que uma IA pode fazer com os dados, precisamos perguntar se os dados estão compreendidos, documentados e autorizados para esse uso.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.

Produto da aula

Ao final da aula, cada grupo deverá entregar:

- uma ficha de documentação do dataset Vozes da Escola ou projeto regional;
- uma tabela com as versões da base;
- uma lista de limitações do dataset;
- uma lista de usos permitidos e usos não recomendados;
- uma breve síntese respondendo:

  "O que uma pessoa precisa saber antes de usar este dataset em uma análise ou em um sistema de IA?"

Avaliação	A avaliação pode ser processual e baseada na participação dos estudantes nas discussões, na análise crítica da base e na qualidade da ficha de documentação produzida. O professor pode observar se os estudantes conseguem explicar a origem dos dados, identificar público respondente e grupos ausentes, reconhecer limitações, definir usos permitidos e usos inadequados, registrar versões da base e apontar cuidados de privacidade. O principal produto avaliativo pode ser composto por: ficha de documentação do dataset, tabela de versionamento, lista de limitações, lista de usos permitidos e não recomendados, e uma breve síntese respondendo à pergunta: "O que uma pessoa precisa saber antes de usar este dataset em uma análise ou em um sistema de IA?".
Material Complementar	• Material Didático do Professor — Tema 308: Definições do Dataset; • Questionário padrão do projeto Vozes da Escola; • Base sintética do projeto Vozes da Escola ou projeto regional; • Base original e base tratada utilizadas nos planos anteriores; • Registro de decisões de limpeza do Plano 307; • Modelo de ficha de documentação do dataset; • Dicionário de dados do projeto; • Lista oficial das iniciativas; • Site do projeto IA@Escola.

CC BY-NC-SA 4.0 VICARI, Rosa Maria; GALAFASSI, Cristiano; BRACKMANN, Christian; MIZUSAKI, Lucas.
IA@Escola. 2024. Disponível em: www.computacional.com.br/ia.