CREDAL: Close Reading of Data Models

George Fletcher
Eindhoven University of Technology
Eindhoven, Netherlands
g.h.l.fletcher@tue.nl

Olha Nahurna
Ukrainian Catholic University
Lviv, Ukraine
onahurna@gmail.com

Matvii Prytula
Ukrainian Catholic University
Lviv, Ukraine
matvii.prytula@ucu.edu.ua

Julia Stoyanovich
New York University
New York, USA
stoyanovich@nyu.edu

ABSTRACT

Data models are foundational to the creation of data and any data-driven system. Every algorithm, ML model, statistical model, and database depends on a data model to function. As such, data models are rich sites for examining the material, social, and political conditions shaping technical systems. Inspired by literary criticism, we propose close readings of data models—treating them as artifacts to be analyzed like texts. This practice highlights the materiality, genealogy, techne, closure, and design of data systems.

While literary theory teaches that no single reading is "correct," systematic guidance is vital—especially for those in computing and data science, where sociopolitical dimensions are often overlooked. To address this gap, we introduce the CREDAL methodology for close readings of data models. We describe its iterative development and share results from a qualitative evaluation, demonstrating its usability and value for critical data studies.

ACM Reference Format:
George Fletcher, Olha Nahurna, Matvii Prytula, and Julia Stoyanovich. 2025. CREDAL: Close Reading of Data Models. In *Workshop on Human-In-the-Loop Data Analytics (HILDA' 25), June 22–27, 2025, Berlin, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3736733.3736737

1 INTRODUCTION

As the saying goes, knowledge is power. Through knowledge, communities make their worlds. Yet, as has been highlighted since Foucault, the converse also holds: power is knowledge [20]. It is primarily those with power who are enabled to resource the data systems that drive contemporary data-driven decision and knowledge making. This, in turn, leads to a perpetuation of the status quo, where what is known is predominantly for and by those with power. Consequently, data systems are too often key enablers and amplifiers of the cruelties of the status quo [3, 5–7, 9, 10, 15, 33, 39, 41, 42]. In this work, we highlight an emerging thread in the broader conversation of how to open up and intervene in cycles of data for more equitable and just futures by focusing on *data models*.

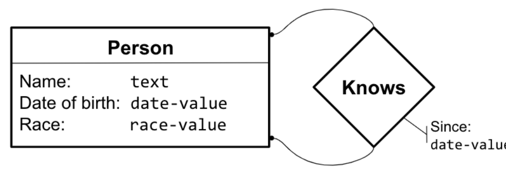

Figure 1: Data model for Example 1.1: `Person` is a class and `Knows` is a relationship between elements of this class. An example instance of this model is: Saori is a `Person` (name "Saori", race "Asian", and date of birth 2001) who `Knows` Kotaro (who is also a `Person`, name "Kotaro", race "Native Hawaiian or Other Pacific Islander", and date of birth 2002) since 2018.

Example 1.1. Consider the data model visualized in Figure 1. Here, "race" takes a single value drawn from a fixed list, standardized "for all Federal reporting purposes" in the USA: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.[1] What does it mean that race is single-valued and drawn from a fixed domain? In which worlds is this meaningful? How does it conflict with other worlds, where race is multi-valued and open-ended? To what ends does the data system even model race in the first place? Why is race an attribute of people, i.e., modeled as inherent (objectively) in a person? Or are people racialized by other people, in which case race would better be modeled as a separate class or as a relationship between people? In what world is race a fixed, limited, single, unmovable, inherent feature of people? Why is the data system provider perpetuating this world? [2]

Conceptual and knowledge modeling of data has been studied in the data management, business information systems, and knowledge representation communities for over 50 years. For overviews of these rich literatures, see the surveys of Veda Storey et al. [2, 40] and Wei Yun et al. [44]. Apart from a rich sub-literature on model validation (i.e., tools and methods to ensure that a model is fit for purpose), we find that the overwhelming majority of work has overlooked broader social aspects of data modeling. Notable exceptions include the qualitative investigation of Graeme Simsion and

HILDA' 25, June 22–27, 2025, Berlin, Germany

ACM ISBN 979-8-4007-1959-2/2025/06
https://doi.org/10.1145/3736733.3736737

[1] https://www.doi.gov/pmb/eeo/directives/race-data
[2] See Chapter 6: The case of race classification and reclassification under apartheid, Bowker and Star [8].

colleagues that highlights the constructive consensus-making nature of modeling in business information systems [37], and recent efforts in the conceptual modeling community on inclusiveness in modeling, see Lukyanenko et al. [27].

How can we systematically unpack a data model, to interpret it as the contextualized creative cultural artifact that it is? Drawing inspiration from literary criticism, our proposal is to closely read data models in the same spirit we read literature. Close reading is a well-known approach to literary analysis that involves careful and sustained interpretation of textual artifacts [21, 24]. Questions to consider during close reading include the speaker, the audiences, the reader's assumed knowledge, the purpose of included details, ambiguous words or phrases, patterns, rhythm, movement, and anything that raises questions or requires clarification.

Students are often taught step-by-step approaches for conducting a close reading of a literary passage.[3] This typically begins with selecting a short section of text and paying attention to unusual or repetitive imagery and themes. The reader is then guided to read the passage, taking notes on language use, repetitions, and changes in tone. The analysis phase involves examining elements such as diction, narrative voice, and rhetorical devices. Following this, a descriptive thesis is formulated based on language observations, and finally, an argument is developed, connecting language use to the broader themes of the text. The method emphasizes the importance of understanding not just how language is used but why, encouraging a thoughtful exploration of the author's intentions and the text's overall significance.

1.1 Why close readings of data models?

When we work with data models, we often treat them as clean abstractions, deliberately setting aside the messy realities they represent. This is, in many ways, the point of modeling: to focus attention on certain aspects of the world while ignoring others. But in doing so, we can lose touch with the broader social, historical, and material contexts behind the data. Close readings of data models offer a way to reconnect with what's been bracketed out. They surface the physical infrastructures where data lives [13, 22, 38], and the interpretive decisions and constraints that shape a model's meaning [4, 12, 18, 25, 26, 28, 31, 35]

Importantly, close readings reveal the *techne* – the practical know-how, tools, and constraints – that shape how models are constructed and used. This challenges the view of modeling as simply documenting reality, and instead positions it as a sociotechnical design process shaped by negotiation, power, and compromise [16, 34, 37]. Close readings prompt critical questions: What was left out of this model, and why? Who benefits from this representation? What futures does it enable—or foreclose?

1.2 Contributions

Although literary theory reminds us there's no single "correct" reading of a text or a model [24], having structured ways to approach these questions is especially important for technical practitioners, who are often trained to ignore the political and ethical dimensions of data work. To our knowledge, a systematic methodology for reading data models currently does not exist.

In this work, we developed CREDAL, a structured methodology for the close reading of data models. Designed to support nuanced exploration of language, relationships, and patterns, CREDAL helps surface biases in modeling choices and promotes the creation of fairer, more just knowledge representations. Beyond its development, our goal was to assess CREDAL 's effectiveness, usefulness, and practical applicability, guided by the following research questions:

RQ1: Is CREDAL usable and useful?
- **RQ1.1:** Is CREDAL easy to learn and apply?
- **RQ1.2:** Does CREDAL improve data modeling proficiency?
- **RQ1.3:** Are learners likely to use CREDAL in the future?

RQ2: Does CREDAL help understand, design and critically evaluate data models?
- **RQ2.1:** Does CREDAL help with understanding data models?
- **RQ2.2:** Does CREDAL alter the approach to structuring and modeling data within modeling tasks?
- **RQ2.3:** Do learners experience a change in their data modeling perspectives after working with CREDAL?

We will present the results of a qualitative study towards answering each of these questions, highlighting the *perceived benefits* of CREDAL, and providing concrete indications for further research and development of the methodology.

2 RELATED WORK

We build on and extend work on *reading data and data models*. Poirier [32] identifies three modes of close reading for datasets: denotative (technical understanding), connotative (cultural context), and deconstructive (representational limits). Through case studies, this work demonstrates how students develop critical insight into the assumptions and politics embedded in datasets. Feinberg [17] complements this with the notion of "slow" data interactions as a reflective alternative to fast, outcome-oriented paradigms. Both emphasize the importance of interpretive engagement with data, but stop short of offering a systematic methodology for close readings, especially one accessible to readers from STEM backgrounds.

Critical data modeling has emerged as a growing area within critical information and data studies. Wickett [43] proposes a formal representation model focusing on the propositional, symbolic, and material aspects of data models, helping surface latent assumptions in data-driven infrastructures. However, this approach largely abstracts away the modelers and those modeled, underemphasizing the relational and interpretive aspects of data. In contrast, Stevens [39] uses autoethnographic methods to explore how modeling choices reflect and perpetuate social power dynamics. Her work positions data models as sites of both domination and possibility, arguing for their potential as tools of intervention and justice.

Our work is also informed by *qualitative and ethnographic studies* that investigate how data professionals engage with modeling in practice. Muller et al. [29] use grounded theory to explore how data scientists construct and navigate data models in real-world settings. Kinnee et al. [23] introduce "autospeculation," an ethnographic method to examine the affective and imaginative dimensions of troubling data models. These studies offer rich insights into the

[3] e.g., UW Madison Writer's Handbook, U. York – Writing at York, UT Austin – Critical Reader's Toolkit

lived experience of data work and highlight the sociotechnical complexities involved in modeling.

Finally, our work aligns with the broader literature on *critical data literacy*, which emphasizes education and public engagement to help students critically examine the social, political, and historical forces shaping data and its uses [11]. We complement this by offering a structured methodology for close readings of data models, supporting deeper reflection and analysis.

3 DEVELOPING THE CREDAL METHODOLOGY

We developed CREDAL through an iterative, stepwise process, detailed below, refining it through practical applications, feedback, and structured evaluation.

3.1 Initial development

3.1.1 Initial Exploration and Refinement. We began by engaging deeply with the technique of close reading, a method from literary studies used for detailed textual analysis [24]. After reviewing relevant literature and practicing on literary texts, each team member independently applied close reading to data schemas, then compared approaches to draft an initial methodology.

Next, we tested the draft on a new set of schemas using a standardized process, which revealed variations in interpretation and helped us refine the method further. We applied the methodology to both knowledge graph schemas (e.g., Wikidata.org, Schema.org) and relational data models. Given the prevalence of relational models in industry and education [36], we focused on them in subsequent rounds. Using open-source data models,[4] we conducted additional close readings and refined CREDAL based on these experiences.

3.1.2 Development of Supplemental Materials. To assess the applicability and effectiveness of the methodology, we involved independent reviewers and created supplemental materials to support their use. Drawing on diverse team perspectives, we developed a practical guide with tips to help users navigate common challenges. We also prepared a sample reading—"A Secure Students' Attendance Monitoring System"[30] (see Appendix)—chosen for its balanced complexity and relatable educational context. These resources offered both conceptual guidance and concrete examples to aid application.

3.2 Iterative refinement

To iteratively refine CREDAL and address our research questions, we developed semi-structured interview questions to gather in-depth feedback from data modeling practitioners. We engaged three increasingly large groups, first introducing the methodology, then collecting their responses.

3.2.1 Initial interviews with undergraduate students. Each team member recruited a volunteer Computer Science student to review CREDAL. Students were given unlimited time to explore the methodology, its materials, and apply it to a provided relational data model in a written format of their choice. After completion, each participated in a 30-minute semi-structured interview, which was audio-recorded, anonymized, and transcribed for analysis. This feedback helped identify areas for improvement, leading to a clearer and more accessible version of the methodology.

3.2.2 A workshop for peer feedback. After refining CREDAL, we held a one-hour workshop with six graduate students in computer and data science at NYU. Participants reviewed the methodology in advance, then collectively applied it to the same relational data model used in earlier stages. We gathered feedback using the same interview questions to ensure consistency and enable comparison across iterations. This approach minimized variability and allowed us to directly assess improvements. Feedback highlighted further refinements, including the recommendation to incorporate more visual aids to improve clarity and ease of use.

3.2.3 Pilot study. To address this feedback, the next group—4 data science graduate students and 7 computer science undergraduates—received a video guide summarizing key points of the methodology with illustrative examples. They were also given an updated methodology guide featuring practical tips and a completed example. Further details about this pilot study are provided in Section 5, with evaluation results discussed in Section 6.

4 A STRUCTURED GUIDE TO CREDAL

We now present a structured guide to CREDAL (Section 4.1) and supporting materials to facilitate its adoption and use (Section 4.2).

4.1 CREDAL

We next present the overall pipeline of CREDAL.

(1) *Define Research Goals and Understand the Data Model.* Set clear objectives for your analysis—model-driven or, if data is available, data-driven. Understand the structure, domain, and notable gaps or omissions in the data model.
(2) *Evaluate Context, Domain Knowledge, and Sources.* (a) Gather domain-specific insights from external sources. (b) Consider ethical and privacy concerns, including omitted sensitive data. (c) Evaluate data sources for bias and incompleteness.
(3) *Exploratory Data Analysis (EDA) and Schema Analysis.* **Data-driven:** Identify patterns, outliers, and anomalies (see [1, 14] for EDA tips). **Model-driven:** Analyze schema structure: (a) Identify entities and relationships; (b) Review constraints (keys, formats, uniqueness); (c) Examine indices and design patterns.
(4) *Explore Related Schemas.* (a) Compare content and structure with related schemas; (b) Note included/excluded elements; (c) Use insights to inform schema improvements.
(5) *Assumptions Loop.* (a) Select a small part of the model (e.g., entity, attribute, or relationship) with potential for bias. (b) Define fairness criteria before identifying negative bias or inequality. (c) **Data-driven:** Quantify potential bias (e.g., skew, missing data). **Model-driven:** Compare related schemas, analyze types/relationships, and conduct domain research. (d) Identify risks and failures: i) Examine how sensitive attributes affect outcomes (where appropriate); ii) Use causal analysis to evaluate fairness; iii) Apply the "5 Whys" to trace root causes (see Section 4.2). (e) Propose mitigation strategies or schema/data enhancements.
(6) *Compile Findings.* Summarize observations in a clear, actionable report that supports future model revision.

[4] e.g., GitMind, DevTools Daily, ConceptDraw

4.2 Supporting Materials

Video tutorial. To improve comprehension, we developed a video tutorial that visually presents key terms and stages of the methodology, with illustrative examples to reinforce each step.

Example of reading. To illustrate the methodology in practice, we developed a step-by-step example using the data model "A Secure Students' Attendance Monitoring System" (see [19] for details). Chosen for its moderate complexity and accessibility, this comprehensive walk-through highlights key nuances and provides clear guidance to support users in applying CREDAL.

Suggestions and best practices. We distilled key techniques that improve understanding and application of the methodology into a set of practical recommendations.

- **Use the 5 Whys method.** This technique helps uncover root causes of bias by repeatedly asking why" to dig deeper into an issue. (a) *Define the problem:* Clearly state the issue to avoid addressing the wrong problem. (b) *Ask the first Why":* Identify the immediate cause of the problem. (c) *Keep asking 'Why':* Continue for each answer until the root cause becomes clear. (d) *Tips:* Move quickly between questions to stay focused, and stop when answers become repetitive or uninformative.
- **Incorporate brainstorming sessions.** Brainstorming fosters creative thinking and diverse perspectives, helping surface hidden biases. Key steps: (a) Clarify the session goal; (b) Record all ideas and discussion; (c) Encourage open, out-loud thinking; (d) Welcome all contributions without judgment; (e) Ask clarifying questions; (f) Organize outcomes for further analysis.
- **Apply principles of literary close reading.** Adapted from literary analysis, close reading offers a structured way to examine data models in depth: (a) Focus on explicit content, avoiding speculation; (b) Read slowly to catch subtle details; (c) Revisit the material multiple times for deeper insight; (d) Analyze each element carefully for accuracy and clarity.

Common pitfalls and considerations. When applying the methodology, users may encounter several potential challenges, summarized below, along with advice for mitigating them:

- **Overlooking entity relationships.** Thoroughly explore the relationships between entities within the schema. Neglecting these connections may lead to incomplete or inaccurate conclusions.
- **Managing Large Data Volumes.** A data-driven approach may introduce complexity when dealing with large datasets. Focus on specific areas of interest and relevant data points to avoid becoming overwhelmed by the volume of information.
- **Bias below the surface.** Bias is not always apparent at the top levels of the schema. Take multiple perspectives, playing the roles of both the creator and reviewer, to identify potential biases or controversies that may lie deeper in the structure.
- **Working with unfamiliar data or schemas.** When dealing with an unfamiliar domain, consider consulting specialists or leveraging external resources to gain a better understanding before proceeding with bias analysis.
- **Curiosity and questioning.** Embrace curiosity and do not hesitate to ask fundamental questions. Seemingly basic inquiries can often lead to important discoveries and unveil hidden aspects of the data model.
- **Concluding the analysis.** After identifying biases, it is essential to clearly articulate your findings and suggest actionable solutions. This ensures that the issues are addressed and mitigated in future iterations of the data model.

5 VALIDATING CREDAL

Involving human participants in the evaluation provides key insights into how well the methodology meets its objectives and promotes critical understanding of data models. Participant engagement helps uncover biases, clarify ambiguities, and validate the content and structure for clarity, conciseness, and relevance—insights only revealed through real-world application and feedback.

To validate CREDAL, we interviewed 11 applied science students—4 from an MS in Data Science and 7 from a BS in Computer Science—at Ukrainian Catholic University in Lviv, Ukraine. Our study received approval from the institution's ethics review board. Since CREDAL targets technical students and practitioners involved in data modeling, analysis, and design, this sample helps assess its relevance and usefulness for its intended audience.

5.1 Interview process

We introduced the CREDAL methodology through a workshop using materials from Section 4, then collected participant feedback via semi-structured interviews. The interview protocol included 14 core questions (see [19] for details), supplemented with follow-up prompts as needed. Questions covered four areas:

- Participants' data modeling background and experience with CREDAL, including general impressions, strengths and weaknesses, and perceived changes in understanding and confidence.
- Feedback on supporting materials and the usefulness of the literary close reading analogy.
- Perceived effectiveness and practical applicability of CREDAL, including likelihood of future use.
- Suggestions for improving CREDAL.

5.2 Interview coding

We recorded and transcribed the interviews, then coded them in the team (see [19] for details). The codebook was developed manually through multiple rounds of independent coding and researcher discussions. We used Atlas.ti to organize the data by linking codes to transcript quotes. Although Atlas.ti supports AI-generated coding aligned with research questions, we chose manual coding to ensure thoroughness and accurately capture emerging themes. To mitigate bias, we independently tagged interview transcripts using two approaches, *questions-based* and *context-based*.

- **Questions-based coding**: We began by reviewing the interview questions and defining corresponding code groups. Based on participants' responses, we identified common themes, created relevant codes, and assigned them to these groups, resulting in 8 groups and 31 codes.
- **Context-based coding**: We analyzed all interviews to identify frequently occurring topics, creating context-based codes and deriving group names accordingly. This approach produced more

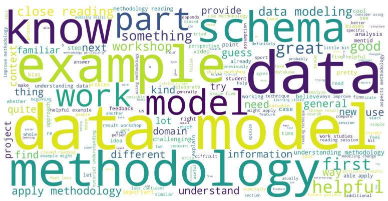

Figure 2: Word cloud of participant responses

fine-grained, specific, and less repetitive codes, resulting in 5 groups and 63 codes.

After applying both approaches, we manually compared their similarities and differences. The authors of each method reviewed overlapping codes—those linked to the same quotations or sharing the same meaning. For code groups, we retained the four that best aligned with our research objectives and questions:

A: Participant background (3 codes)
B: Methodology effectiveness (3 codes), for research questions **RQ1.2**, **RQ1.3**, **RQ2.1**, **RQ2.2**, and **RQ2.3**
C: Methodology improvement (6 codes)
D: Methodology strengths (3 codes), for research question **RQ1.1**

See [19] for a list of codes, code groups, and explanations.

6 RESULTS ANALYSIS AND DISCUSSION

All 11 participants had prior data modeling experience: 2 at a basic level, 7 intermediate, and 2 advanced. None were familiar with the literary close reading methodology before the workshop. Despite individual perspectives, common themes emerged across interviews. We synthesized key patterns, which are discussed throughout this section. As a starting point, Figure 2 presents a word cloud of participant responses, highlighting frequent terms such as "methodology", "data", "data model", "know", and "example".

Figure 3 summarizes code frequencies from Section 5.2, with colors indicating the four code groups. The most frequent group was *methodology strengths* (e.g., "Helpful supplemental materials", "Helpful structure", "Encourages analysis and reflection"), followed by *methodology effectiveness* (e.g., "Likely to use in future", "Improved proficiency"), and then *methodology improvement* (e.g., "Improve presentation of bias", "Improve supplemental materials"). These results suggest that participants found CREDAL helpful and effective, while also highlighting the need to improve the supporting materials more than the methodology itself.

6.1 RQ1: Usability and Usefulness of CREDAL

The first research question **RQ1: Is CREDAL usable and useful?** comprises three subquestions.

6.1.1 ***RQ1.1****: Is* CREDAL *easy to learn and apply?* Participants highlighted the methodology's "Helpful Structure" (13 entries) and especially valued the "Helpful Supplemental Materials" (28 entries—the most frequent code). These suggest the approach is feasible and well-supported by examples and resources that aid implementation. Representative quotes include:

> [P10]: "I would say that the steps were defined pretty clearly. And also, there was a pretty helpful example, which went really in-depth, making it easy to work on my schema after reading the entire example with a different schema."

> [P5]: "I read that guide tutorial that was sent to me; it was very detailed and contained much information about how to perform this close reading."

6.1.2 ***RQ1.2****: Does* CREDAL *improve data modeling proficiency?* The code "Improved participant's data modeling proficiency" appeared 12 times, ranking among the top five. It reflects feedback that the workshop enhanced participants' conceptual and practical understanding—even among those with prior experience. Participants noted improvements with comments such as:

> [P5]: "I think that understanding changed, and I can also say that I have a new view on data models after reading that tutorial."

> [P9]: "I feel more confident; gaining new knowledge about data modeling has strengthened my skills."

6.1.3 ***RQ1.3****: Are learners likely to use* CREDAL *in the future?* The code "Participant likely to use methodology in the future" appears 18 times, indicating strong interest in applying CREDAL to future projects. Participants often shared specific use cases, such as:

> [P8]: "Enhancing the training data, especially for R&D projects, I guess that might be a good push and a good idea to read the work that you've accomplished."

> [P4]: "A schematic framework that could help illustrate why the data model is good or bad."

Based on these results, we conclude that CREDAL was found to be usable and useful by participants.

6.2 RQ2: CREDAL helps work with data models

The second research question **RQ2: Does CREDAL help understand, design, and critically evaluate data models?** focuses on the practical application of the CREDAL and its interaction with real data models, leading to three subquestions.

6.2.1 ***RQ2.1****: Does* CREDAL *help with understanding data models?* The code "Helped participant understand a specific model" reflects increased confidence and improved data modeling skills reported by most participants after using CREDAL. For example:

> [P11]: "It refreshed my knowledge about data modeling and also taught me some new ways to work with them, especially to analyze data models more carefully and look for potential problems."

> [P7]: "I read that tutorial and learned how to perform close reading on those data models."

6.2.2 ***RQ2.2****: Does* CREDAL *alter the approach to structuring and modeling data within modeling tasks?* Participants shared that, before the workshop, they rarely considered how data models were constructed. Afterward, many reported a shift in perspective and a readiness to refine their modeling approach—reflected in quotes coded as "Improved participant's data modeling proficiency":

> [P8]: "Definitely, I guess I've almost never taken care about where the data was sourced and why someone decided to model it in a certain way."

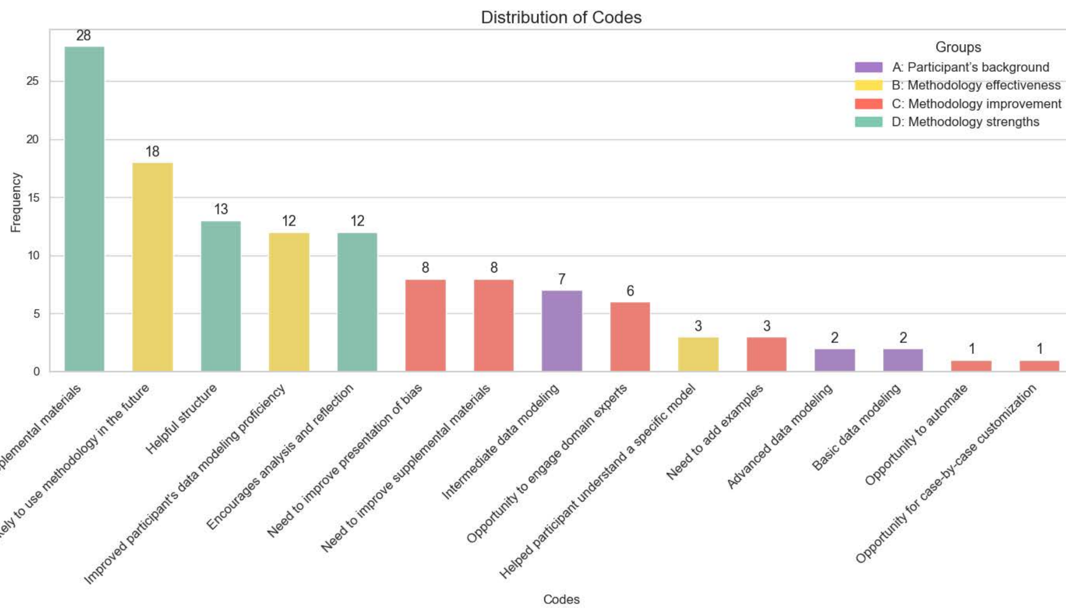

Figure 3: Interviews codes frequency distribution

> [P11]: "The workshop provided me with a structured way to evaluate data models. It helped me spot issues before they became problems."

Most participants had intermediate data modeling experience, yet even those with prior knowledge reported gaining new insights relevant to their studies or work. These reflections are captured under the code "Improved participant's data modeling proficiency." Representative examples include:

> [P3]: "You can identify where you might be wrong about your granularity or your assumptions about whether a data model could work here or not. Without some methodology, you might miss those crucial aspects, even if you have experience working with other models."

> [P1]: "It's essential to ensure that biases don't negatively impact the model's performance."

Our results suggest that CREDAL supports data modeling work. Participants reported improved analysis and evaluation skills, and proposed new design approaches, demonstrating CREDAL 's effectiveness in fostering deeper understanding and application.

6.3 Opportunities for Improvement

Participants provided valuable feedback on enhancing CREDAL 's clarity, usability, and adaptability. We summarize their feedback here, see [19] for additional information.

A key theme was the need for clearer, more structured supplemental materials. As one participant noted, *"The whole methodology can be expressed in a more concise and step-by-step approach."* Others emphasized incorporating visual aids and additional examples to better support understanding and identify biases. Bias detection was highlighted as particularly challenging, with calls for clearer definitions and guidance due to its subjective nature.

Suggestions for extending the methodology included engaging domain experts and automating repetitive tasks to improve efficiency and scalability. One participant remarked, *"Doing it manually is a little bit over-engineering. I would suggest trying to automate this methodology."* Others recommended offering flexible versions of CREDAL —a comprehensive version for complex projects and a streamlined one for simpler tasks. These insights point to promising directions for making CREDAL more accessible, modular, and effective across diverse data modeling contexts.

7 CONCLUSIONS AND FUTURE WORK

Close readings of data models reconnect us with their materiality, origins, design, and constraints. We introduced CREDAL, the first systematic methodology for this practice—along with qualitative findings demonstrating its usability, usefulness, and effectiveness for critical data analysis. CREDAL was developed and evaluated with undergrad and grad students, mainly focused on relational data models. Future work could expand to broader audiences and non-relational paradigms (e.g., knowledge graphs, ontologies). (2) Interviews revealed areas for improvement, including enhanced supplemental materials, better support for domain expert collaboration, customizable workflows, and automation of repetitive tasks. These are key opportunities for future research and development.

Acknowledgments. This research was conducted as part of the RAI for Ukraine program, run by the Center for Responsible AI at New York University in collaboration with Ukrainian Catholic University in Lviv. This work was supported in part by a grant from the Simons Foundation (SFARI award #1280457, JS), and by NSF Awards No. 2326193 and 1922658.

REFERENCES

[1] Ziawasch Abedjan, Lukasz Golab, Felix Naumann, and Thorsten Papenbrock. 2018. *Data profiling.* Morgan & Claypool Publishers.

[2] Jacky Akoka, Isabelle Comyn-Wattiau, Nicolas Prat, and Veda C. Storey. 2024. Unraveling the foundations and the evolution of conceptual modeling—Intellectual structure, current themes, and trajectories. *Data & Knowledge Engineering* 154 (2024), 102351.

[3] Josef Ansorge. 2016. *Identify and sort.* Hurst and Co.

[4] Ariella Aïsha Azoulay. 2019. *Potential history: Unlearning imperialism.* Verso.

[5] Christoph Becker. 2023. *Insolvent.* MIT Press.

[6] Ruha Benjamin. 2019. *Race after technology.* Polity.

[7] Katherine Bode and Lauren M. E. Goodlad. 2023. Data worlds: An introduction. *Critical AI* 1, 1-2 (2023).

[8] Geoffrey C. Bowker and Susan Leigh Star. 1999. *Sorting things out.* MIT Press.

[9] Sasha Costanza-Chock. 2020. *Design justice.* MIT Press.

[10] Nick Couldry and Ulises A. Mejias. 2019. *The costs of connection.* Stanford University Press.

[11] Aayushi Dangol and Sayamindu Dasgupta. 2023. Constructionist approaches to critical data literacy: A review. In *ACM IDC.* 112–123.

[12] Mimi Ọnụọha. 2016. On missing data sets. https://github.com/MimiOnuoha/missing-datasets.

[13] Paul Dourish. 2017. *The stuff of bits.* MIT Press.

[14] Allen Downey. 2014. *Think Stats: Exploratory data analysis, 2nd Ed.* O'Reilly.

[15] Virginia Eubanks. 2018. *Automating inequality.* St Martin's Press.

[16] Melanie Feinberg. 2017. A design perspective on data. In *CHI.* ACM, 2952–2963.

[17] Melanie Feinberg. 2017. Reading databases: slow information interactions beyond the retrieval paradigm. *Journal of Documentation* 73, 2 (2017), 336–356.

[18] Melanie Feinberg. 2022. *Everday adventures with unruly data.* MIT Press.

[19] George Fletcher, Olha Nahurna, Matvii Prytula, and Julia Stoyanovich. 2025. CREDAL: Close Reading of Data Models. *arXiv:2502.07943 [cs.DB]* (2025).

[20] Michel Foucault. 1980. *Power/Knowledge: Selected interviews and other writings 1972-1977, C. Gordon (Ed.).* Pantheon Books.

[21] John Guillory and Scott Newstok. 2025. *On Close Reading.* University of Chicago Press.

[22] Mél Hogan. 2024. The fumes of AI. *Critical AI* 2, 1 (04 2024).

[23] Brian Kinnee, Audrey Desjardins, and Daniela Rosner. 2023. Autospeculation: Reflecting on the intimate and imaginative capacities of data analysis. In *ACM CHI.* Article 151, 10 pages.

[24] Frank Lentricchia and Andrew DuBois (Eds.). 2003. *Close reading: The reader.* Duke UP.

[25] Sabina Leonelli. 2015. What counts as scientific data? A relational framework. *Philosophy of Science* 82, 5 (2015), 810–821.

[26] Sabina Leonelli and Niccolò Tempini (Eds.). 2020. *Data journeys in the sciences.* Springer.

[27] Roman Lukyanenko, Dominik Bork, Veda C. Storey, Jeffrey Parsons, and Oscar Pastor. 2023. Inclusive conceptual modeling: Diversity, equity, involvement, and belonging in conceptual modeling. In *ER Forum.*

[28] Nicolas Malevé. 2021. On the data set's ruins. *AI and Society* 36 (2021), 1117–1131.

[29] Michael Muller, Ingrid Lange, Dakuo Wang, David Piorkowski, Jason Tsay, Q. Vera Liao, Casey Dugan, and Thomas Erickson. 2019. How data science workers work with data: Discovery, capture, curation, design, creation. In *ACM CHI.* 1–15.

[30] Oyetola Oluwadamilola, Ayodeji Okubanjo, and Olawale Olaluwoye. 2017. A Secure Students Attendance Monitoring System. *Journal of Engineering and Technology* 2, 2 (2017).

[31] Allison Parrish. 2016. Programming is forgetting. http://opentranscripts.org/transcript/programming-forgetting-new-hacker-ethic/. Keynote, Open Hardware Summit 2016.

[32] Lindsay Poirier. 2021. Reading datasets: Strategies for interpreting the politics of data signification. *Big Data & Society* 8, 2 (2021).

[33] Holly Randell-Moon and Ryan Tippet (Eds.). 2016. *Security, race, biopower: Essays on technology and corporeality.* Palgrave Macmillan.

[34] Ryan Shaw. 2023. Conceptual modeling as language design. *J Assoc Inf Sci Technol* (2023).

[35] Jihan Sherman, Romi Morrison, Lauren Klein, and Daniela Rosner. 2024. The power of absence: Thinking with archival theory in algorithmic design. In *ACM DIS.* 214–223.

[36] Avi Silberschatz, Henry F. Korth, and S. Sudarshan. 2019. *Database System Concepts, Seventh Edition.* McGraw-Hill.

[37] Graeme Simsion, Simon K. Milton, and Graeme Shanks. 2012. Data modeling: Description or design? *Information & Management* 49, 3 (2012), 151–163.

[38] Brian Cantwell Smith. 1985. The limits of correctness. *SIGCAS Comput. Soc.* 14-15, 4, 1-3 (1985), 18–26.

[39] Nikki Lane Stevens. 2022. *Modeling power: data models and the production of social inequality.* Ph. D. Dissertation. Arizona State University.

[40] Veda C. Storey, Roman Lukyanenko, and Arturo Castellanos. 2023. Conceptual modeling: Topics, themes, and technology trends. *ACM Comput. Surv.* 55, Article 317 (2023).

[41] Kai von Lewinski, Michael Beurskens, and Stefanie Scherzinger. 2024. Data modelling as a means of power: At the legal and computer science crossroads. *Comput. Law Secur. Rev.* 52 (2024), 105865.

[42] Joseph Weizenbaum. 1976. *Computer power and human reason.* W. H. Freeman and Co.

[43] Karen M. Wickett. 2023. Critical data modeling and the basic representation model. *J Assoc Inf Sci Technol* (2023).

[44] Wei Yun, Xuan Zhang, Zhudong Li, Hui Liu, and Mengting Han. 2021. Knowledge modeling: A survey of processes and techniques. *Int. J. Intell. Syst.* 36, 4 (2021), 1686–1720.