MOTIVATION	
For what purpose was the dataset created?	The NoRa dataset was created to evaluate the robustness of large language models when faced with noisy rationales in chain-of-thought prompting. It addresses the under-explored challenge of LLMs' reasoning ability when presented with in-context examples containing irrelevant or inaccurate steps.
Who created the dataset and on behalf of which entity?	The authors created the dataset. Keep anonymous in the submission phrase.
Who funded the creation of the dataset?	Unavailable.
Any other comment?	The NoRa dataset is the first of its kind specifically designed to evaluate the robustness of LLMs against noisy rationales. It covers a wide range of reasoning tasks, including mathematical, symbolic, and commonsense reasoning, The dataset includes varying levels of noise in rationales to systematically access LLMs' performance.
COMPOSITION	
What do the instances that comprise the dataset represent?	Each instance in the NoRa dataset consists of a question, its correct answer, and in-context demonstrations containing noisy reasoning steps.
How many instances are there in total?	The NoRa dataset contains 26,391 questions in total (see Section 3.2). Table 2 contains the sizes of our datasets.
Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?	The NoRa dataset is a carefully curated sample selected from a larger pool of potential reasoning tasks.
What data does each instance consist of?	Each instance in the NoRa dataset contains a question, its correct answer, and in-context demonstrations with noisy reasoning steps.
Is there a label or target associated with each instance?	Yes, each instance has a correct answer that serves as its label.
Is any information missing from individual instances?	No, each instance contains complete information.
Are relationships between individual instances made explicit?	Yes, instances are grouped by reasoning type: mathematical, symbolic, and commonsense.
Are there recommended data splits?	No, this dataset is designed for training-free in-context learning. We provide recommended in-context demonstrations for each question.
Are there any errors, sources of noise, or redundancies in the dataset?	No, the dataset itself is clean and does not contain any errors, unintended noise, or redundancies.
Is the dataset self-contained, or does it link to or otherwise rely on external resources?	Yes, the dataset is self-contained and does not rely on any external resources.
Does the dataset contain data that might be considered confidential?	No.
Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?	No, the dataset does not include any offensive, insulting, threatening, or anxiety-inducing content.
COLLECTION	
How was the data associated with each instance acquired?	The data was created based on the reference of three rational datasets. (see sec 3.2)
What mechanisms or procedures were used to collect the data?	We systematically added irrelevant or inaccurate thoughts to clean rationales from existing datasets. The irrelevant or inaccurate thoughts are generated from GPT4.
If the dataset is a sample from a larger set, what was the sampling strategy?	Whenever we use subsets, we randomly sample from the original data.
Who was involved in the data collection process and how were they compensated?	The authors of the submission.
Over what timeframe was the data collected?	NoRa was created in 2023. The original datasets of NoRa (Base Calculation, SCAN, and CLUTRR) were created between 2018-2023. We added noisy rationales to these existing datasets to create NoRa.
Were any ethical review processes conducted?	No, as the dataset does not involve human subjects or sensitive information, no formal ethical review was required. However, we adhered to ethical guidelines for AI research throughout the process
PREPROCESSING	
Was any preprocessing/cleaning/labeling of the data done?	Yes. We preprocess the original datasets by adding noisy rationales (irrelevant or inaccurate thoughts) generated by GPT-4 and ensure that the questions and answers remained unchanged. (detailed in Section 3.1)
Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data?	Yes. We retained the original datasets (Base Calculation, SCAN, and CLUTRR) in their raw form, as well as our modified NoRa dataset with added noisy rationales.
Is the software that was used to preprocess/clean/label the data available?	No.
USES	
Has the dataset been used for any tasks already?	Yes, the NoRa dataset has been used in our paper to benchmark the reasoning robustness of large language models against noisy rationales in chain-of-thought prompting.
Is there a repository that links to any or all papers or systems that use the dataset?	Keep anonymous in the submission phrase.
What (other) tasks could the dataset be used for?	The NoRa dataset could be used for: 1. Evaluating and improving LLMs' robustness to noisy information 2. Developing new denoising techniques for reasoning tasks 3. Understanding the impact of In-context noise on LLMs.
Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?	The dataset contains artificially introduced noise, which might not perfectly reflect real-world scenarios of noisy reasoning.
Are there tasks for which the dataset should not be used?	The dataset should not be used for tasks that require completely noise-free data or for evaluating models' performance on clean, standard reasoning tasks without noise.
DISTRIBUTION	
Will the dataset be distributed to third parties outside of the entity on behalf of which the dataset was created?	Yes, we plan to make the dataset available to the research community.
How will the dataset will be distributed?	The dataset will be made available through a public repository (e.g., GitHub) along with necessary documentation for its use.
When will the dataset be distributed?	We aim to release the dataset upon publication of our paper.
Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?	Not to our knowledge.
Have any third parties imposed IP-based or other restrictions on the data associated with the instances?	Not to our knowledge.
Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?	Not to our knowledge.
MAINTENANCE	
Who will be supporting/hosting/maintaining the dataset?	Not available. Keep anonymous in the submission phrase.
How can the owner/curator/manager of the dataset be contacted?	Not available. Keep anonymous in the submission phrase.
Is there an erratum?	No.
Will the dataset be updated?	No.
If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?	No.

Table 1: Datasheet for the NoRa dataset, following the framework introduced by gebru2021datasheets.

Clean Example
Inputs: Question: In base-9, what is 86+57? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". Rationale: In base-9, the digits are "012345678". We have 6 + 7 = 13 in base-10. Since we're in base-9, that exceeds the maximum value of 8 for a single digit. 13 mod 9 = 4, so the digit is 4 and the carry is 1. We have 8 + 5 + 1 = 14 in base 10. 14 mod 9 = 5, so the digit is 5 and the carry is 1. A leading digit is 1. So the answer is 154. Answer: Answer:\box{154}. Question: In base-9, what is 63+34? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". **Output:** Rationale: In base-9, the digits are "012345678". We have 3 + 4 = 7 in base-10. Since we're in base-9, 7 doesn't exceed the maximum value of 8 for a single digit. 7 mod 9 = 7, so the digit is 7 and the carry is 0. We have 6 + 3 + 0 = 9 in base 10. 9 mod 9 = 0, so the digit is 0 and the carry is 1. A leading digit is 1. So the answer is 107. Answer: Answer:\boxed{107}.
Irrelevant Noisy Example
Inputs: Question: In base-9, what is 64+62? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". Rationale: In base-9, the digits are Ö12345678·. We have 4 + 2 = 6 in base-10. In base-9, the addition of 4 and 2 may result in a carry to the next place value, depending on the sum of the digits. Since we're in base-9, that doesn't exceed the maximum value of 8 for a single digit. When adding numbers in base-9, it's crucial to recognize when a sum exceeds the maximum value allowed for a single digit, as this indicates the need for a carry. 6 mod 9 = 6, so the digit is 6 and the carry is 0. We have 6 + 6 + 0 = 12 in base 10. 12 mod 9 = 3, so the digit is 3 and the carry is 1. A leading digit is 1. The value of the leading digit in a number contributes significantly to the overall magnitude of the number, as it is multiplied by the highest power of the base in the place value system. So the answer is 136. In base-9, each digit in the answer 136 contributes to the overall value of the number based on its place value and the powers of the base. Answer: Answer:\box{136}. Question: In base-9, what is 86+57? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". **Output:** Rationale: In base-9, the digits are "012345678". We have 6 + 7 = 13 in base-10. Since we're in base-9, the maximum value for a single digit is 8, so we need to consider the carry to the next place value. 13 mod 9 = 4, so the digit is 4 and the carry is 1. We have 8 + 5 + 1 = 14 in base 10. 14 mod 9 = 5, so the digit is 5 and the carry is 1. A leading digit is 1. The value of the leading digit in a number contributes significantly to the overall magnitude of the number, as it is multiplied by the highest power of the base in the place value system. So the answer is 150. In base-9, each digit in the answer 150 contributes to the overall value of the number based on its place value and the powers of the base. Answer: Answer:\boxed{150}.
Inaccurate Noisy Example
Inputs: Question: In base-9, what is 52+55? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". Rationale: In base-9, the digits are "012345678". We have 2 + 5 = 7 in base-10. Since we're in base-9, that doesn't exceed the maximum value of 8 for a single digit. 7 mod 9 = 7, so the digit is 7 and the carry is 0. We have 5 + 5 + 0 = 10 in base 10. 10 + 3 = 13. 10 mod 9 = 1, so the digit is 1 and the carry is 1. A leading digit is 1. 1 + 9 = 10. So the answer is 117. Answer: Answer:\box{117}. Question: In base-9, what is 81+31? Please reason it step by step. End the response with the result in "Answer:\boxed{result}". **Output:** Rationale: In base-9, the digits are "012345678". We have 1 + 1 = 2 in base-10. Since we're in base-9, this is less than 9, so there is no carry. Therefore, the digit is 2. We have 8 + 3 = 11 in base-10. Since 11 exceeds the maximum value of 8 for a single digit in base-9, we have a carry of 1. The sum of 8 and 3 is 11, so our digit is 2 and we have a carry of 1. Considering the carry, 1 + 1 = 2. Therefore, the digit in the first place of the sum is 2 and there is no more carry. A leading digit is 0. So the answer is 202. Answer: Answer:\boxed{202}.

Table 2: Performances with different demonstrations.

Target Rationale	
In base-9, the digits are "012345678". We have 3 + 8 = 11 in base-10. Since wefe in base-9, that exceeds the maximum value of 8 for a single digit. 11 mod 9 = 2, so the digit is 2 and the carry is 1. We have 6 + 1 + 1 = 8 in base 10. 8 mod 9 = 8, so the digit is 8 and the carry is 0. A leading digit is 0. So the answer is 82. Answer:\boxed{82}.	
Irrelevant Hard Noisy Rationale	Inaccurate Hard Noisy Rationale
n base-9, the digits are "012345678". We have 3 + 8 = 11 in base-10. In physics, ... Since we're in base-9, that exceeds the maximum value of 8 for a single digit.In computer science, ... We have 6 + 1 + 1 = 8 in base 10. In mathematics, ... 8 mod 9 = 8, so the digit is 8 and the carry is 0. A leading digit is 0. In game theory, ... So the answer is 82. In chemistry, ... Answer:\box{82}	In base-9, the digits are "012345678". 9 + 1 = 10. We have 3 + 8 = 11 in base-10. 11 + 6 = 17. Since wefe in base-9, that exceeds the maximum value of 8 for a single digit.8 + 6 = 14. 11 mod 9 = 2, so the digit is 2 and the carry is 1. We have 6 + 1 + 1 = 8 in base 10. 8 + 3 = 11. 8 mod 9 = 8, so the digit is 8 and the carry is 0. 8 + 9 = 17. A leading digit is 0. 0 + 9 = 9. So the answer is 82. Answer:\box{82}.
Denoised Rationale by CD-CoT	Denoised Rationale by CD-CoT
n base-9, the digits are "012345678". We have 3 + 8 = 11 in base-10. Since we're in base-9, that exceeds the maximum value of 8 for a single digit. 11 mod 9 = 2, so the digit is 2 and the carry is 1. We have 6 + 1 + 1 = 8 in base 10. 8 mod 9 = 8, so the digit is 8 and the carry is 0. A leading digit is 0. So the answer is 82. Answer:\boxed{82}.	In base-9, the digits are "012345678". We have 6 + 8 = 14 in base-10. Since we're in base-9, that exceeds the maximum value of 8 for a single digit. 14 mod 9 = 5, so the digit is 5 and the carry is 1. We have 3 + 1 + 1 = 5 in base 10. 5 mod 9 = 5, so the digit is 5 and the carry is 0. A leading digit is 0. So the answer is 50. Answer:\boxed{50}.

Table 3: Rephrased Rationale in Different Noisy Settings.