Appendices

A Specialized Prompt to LLM

In the following table, we present a breakdown of the prompt given to the chosen LLMs for our data-specific tasks. The LLM is provided with exact details for the requirements in format, word length, and categories (if any).

Table 4: Tabular Representation of Prompt for Specialized Feature Extraction

Request	Description
Objective	*”I need to build a structured dataset. I will provide individual cases, and you will extract and infer specific features, ensuring consistency in formatting. The dataset should be output in CSV format.”*
Dataset Structure	Case Summary (text, max 50 words) – A concise summary of the case. Action (text, max 3 words) – The key action taken by the active agent. Domain (categorical, select from predefined list) – The ethical domain the case falls under. Active Agent (text, max 3 words) – The person/entity responsible for the action. Passive Agent (text, max 3 words) – The person/entity affected by the action. Consequence (text, max 3 words) – The main outcome resulting from the action. Severity of Consequence (float, -1 to +1) – How significant the consequence is. Utility of Consequence (float, -1 to +1) – How beneficial or harmful the consequence is. Duration of Consequence (float, -1 to +1) – The long-term impact of the consequence. Moral Intention of Active Agent (float, -1 to +1) – The intent behind the action. Ethical Principles Upheld or Violated (text, max 5 words, float, -1 to +1) – The ethical principles affected and their aggregated score. Moral Decision (categorical: ’morally right’, ’morally wrong’, ’morally grey’) – The moral judgment of the action. Moral Decision Explanation (text, max 50 words) – A brief justification for the decision.
Feature Definitions and Constraints	Numerical Scores (-1 to +1): Utility of Consequence: -1 (extremely bad) to +1 (extremely good). Severity of Consequence: -1 (extreme bad severity) to +1 (extreme good severity). This score aligns with Utility of Consequence. Duration of Consequence: -1 (extreme long-term bad) to +1 (extreme long-term good). This score aligns with Utility of Consequence. Moral Intention: -1 (bad intent) to +1 (good intent), independent of the consequence. Ethical Principles Score: The average of all principles upheld (+1) or violated (-1).
Ethical Principles Consideration	The following ranking applies when assigning ethical principle scores: Integrity > 2. Respect > 3. Reciprocity > 4. Accountability > 5. Financial Competence. The dataset must reflect that human life > animal life > material goods and ecosystem preservation > monetary loss when assigning scores.
Predefined Domain Categories	(Select the most relevant one per case) Bioethics, Medical Ethics, Research Ethics, Animal Ethics, Corporate Ethics, Environmental Ethics, AI & Data Ethics, Neuroethics Tech & Cyber Ethics, Legal Ethics, Media & Journalism Ethics, Military Ethics, Political Ethics, Sports Ethics, Educational Ethics, Sexual Ethics, Religious Ethics
Processing Guidelines	Each case should be evaluated independently – no case should influence another. Textual fields should be concise (max 3 words except for summaries and explanations). Scoring must align with predefined moral considerations and rankings. Final moral decision should be based on moral intent, consequence severity, and ethical violations/upholdings.
Case	< Extracted raw selftext from Reddit>

B Questionnaire for Expert Perception of LLM Ethical Cognition

The following table presents the structure of the questionnaire given to the team of 8 expert ethicists associated with this work. It gives two kinds of questions; type D, for demographic type questions to gauge some details of the respondents, and Q for the main questions of the questionnaire. We have reported the average ratings of the responses for each type of question. Note that the Q-type questions represent only one sample case, however, the ethicists were provided a collection of different sample cases requiring similar ratings.

Table 5: Questionnaire on domain expert perception of LLM moral cognition for one sample case.

Code	Question	Avg. Rating
D1	On a scale of 1-5 (where 1 is lowest and 5 is highest), how would you rate your expertise in the study of ethics?	3.63
D2	How would you rate your expertise in the study of philosophy and related subjects?	3.25
D3	How would you rate your understanding on LLMs and how they work?	3.00
D4	How would you rate your confidence in the answers given by an LLM for objective-type questions?	3.50
D5	How would you rate your confidence in the answers given by an LLM for subjective-type questions?	2.63
Q1	The LLM provided us with the following summary based on our given prompt (max 50 words): ”Sammy and his twin daughters moved in after his divorce, but they repeatedly took Zoey’s belongings without permission. After Sloane ruined Zoey’s expensive makeup, her father dismissed concerns. The user installed a lock, angering Sammy and his wife. Tensions rose, leading to silent treatment and conflict within the household.”Rate the summary on a scale of 1 to 5.	4.19
Q2	The LLM inferred and extracted the following main action (max 5 words) from the above case: ”Installed lock on Zoey’s door”Rate this response on a scale of 1 to 5.	4.03
Q3	The LLM inferred and extracted the following area of applied ethics (max 5 words) from the above case: ”Personal privacy and property”Rate this response on a scale of 1 to 5.	4.45
Q4	The LLM inferred and extracted the following Active Agent (max 5 words) from the above case: ”Father (narrator)”Rate this response on a scale of 1 to 5.	4.52
Q5	The LLM inferred and extracted the following Passive Agent (max 5 words) from the above case: ”Zoey, Sammy’s daughters”Rate this response on a scale of 1 to 5.	4.29
Q6	The LLM inferred and extracted the following consequence (max 5 words) from the above case: ”Family conflict and silent treatment”Rate this response on a scale of 1 to 5.	3.90
Q7	The LLM inferred and extracted the following severity of the consequence from the above case: ”Moderate Severity”. (From categories mild, moderate, significant).Rate this response on a scale of 1 to 5.	4.06
Q8	The LLM inferred and extracted the following Utility of Consequence from the above case: ”Good”. (From categories good, bad, grey).Rate this response on a scale of 1 to 5.	3.48
Q9	The LLM inferred and extracted the following Duration of Consequence from the above case: ”Short-term” (From categories short-term and long-term).Rate this response on a scale of 1 to 5.	4.03
Q10	The LLM inferred and extracted the following Moral intention of the active agent from the above case: ”Good” (From categories good, bad, grey).Rate this response on a scale of 1 to 5.	3.87
Q11	The LLM inferred and extracted the following ethical principles upheld from the above case: ”Privacy, autonomy, fairness”.Rate this response on a scale of 1 to 5.	4.19
Q12	The LLM inferred and extracted the following ethical principles violated from the above case: ”Beneficence”.Rate this response on a scale of 1 to 5.	3.29
Q13	The LLM inferred and extracted the following moral decision from the above case: ”Morally right”.Rate this response on a scale of 1 to 5.	3.61
D6	How would you now rate your confidence in the answers given by an LLM for objective-type questions?	3.38
D7	How would you now rate your confidence in the answers given by an LLM for subjective-type questions?	3.50