Table 4: Conversations quality human evaluation results (1–5 scale). Higher is better.

Chat Size	Coherence and Flow	Dialogue Realism	Complexity and Depth
128K	4.4	4.55	4.35
500K	4.49	4.4	4.63
1M	4.66	4.54	4.6
10M	4.6	4.8	5
Average	4.53	4.57	4.64

B.3 BENCHMARK CREATION DETAILS

B.3.1 DOMAIN COVERAGE OF THE DATASET

To ensure broad coverage and realism, our dataset spans a diverse set of domains. The collection includes both technical and non-technical conversations, ranging from specialized domains such as coding, mathematics, financial investment and health to personal and social domains such as therapy, lifestyle, and trip planning. In total, we designed 100 multi-turn chats distributed across 19 domains, each represented by a set of distinct titles that capture the thematic scope of the dialogues. The full list of domains and their associated chat titles is provided in Table 5.

Table 5: Domains and associated chat titles in our dataset (100 total chats).

Domain	Chat Titles
Coding	Designing a Large-Scale Retrieval-Augmented Generation (RAG) System for Enterprise Search • Creating a Self-Driving Car Simulation Environment • Developing a Multi-Agent AI Research Platform • Building a Multi-Language AI Chatbot with Contextual Memory • Developing a Personalized News Aggregator with AI Summarization • Creating an Autonomous Stock Trading Bot • Implementing a Custom Image Captioning Model • Building a Multiplayer Online Game with Real-Time Physics • Building a Real-Time Chat Application with Node.js and Socket.io • Creating an AI-Powered Resume Analyzer with Python and NLP • Developing a Computer Vision App for Real-Time Object Detection • Creating a Restaurant Recommendation System • Automating Social Media Posts with Python • Building a Personal Budget Tracker Web App in Python and Flask • Creating a Command-Line To-Do List Manager in Go • Developing a Weather Forecast App in JavaScript with OpenWeather API • Training a Spam Email Classifier Using Python and Scikit-learn • Building a Portfolio Website with HTML, CSS, and Bootstrap
Math	Partial Differential Equations (PDEs) in Depth • Functional Analysis and Infinite-Dimensional Spaces • Solving Ordinary Differential Equations (ODEs) • Deep Dive into Number Theory • Advanced Probability and Combinatorics • Exploring Non-Euclidean Geometry • Studying Multivariable Calculus • Diving into Analytic Geometry • Developing Skills in Mathematical Induction • Exploring Conic Sections in Depth • Understanding Sequences and Series • Mastering Basic Differential Calculus • Exploring the Geometry of Triangles • Understanding the Basics of Probability • Mastering Algebraic Equations for Everyday Problem Solving • Learning the Foundations of Trigonometry • Mastering Fractions, Decimals, and Percentages
Writing Assistant & Learning	Building a Portfolio-Ready Resume that Passes Any Applicant Tracking System • Mastering the Art of Persuasive Academic Essay Writing • Crafting a Standout Cover Letter for Competitive Job Markets • Designing a Multi-Purpose Personal Statement for Global Opportunities • Developing a Self-Editing System for Lifelong Writing Improvement