Workstream	Task	Description	Owner(s)
Simulator	Dual-Agent Setup	Develop the "Dual-Agent" pipeline where one LLM simulates a "Distressed Client" using DSM-5 personas and another plays a "CBT Therapist".	Yujun, Yifan, Akshaya
Simulator	Generate synthetic preference pairs	Setup Distilabel (for high-throughput generation) or DataDreamer (for prompt chaining) to automate clinical scenarios and critiques.	Yujun, Yifan, Akshaya
Reward construction	Affective Mapping	Implement the GoEmotions valence mapping to transform discrete emotion probabilities into a continuous scale from [-1, 1]	Maisy, Yuze
Reward construction	Map GoEmotions labels to the Hourglass of Emotions framework	Map GoEmotions labels to the Hourglass of Emotions' four dimensions: Introspection, Temper, Attitude, and Sensitivity	Maisy, Yuze
Reward construction	Unified Affective Matrix Reward Engine construction	Build the function to calculate standardized polarity in the format of the Unified Affective Matrix Reward Engine, which serves as the primary reward signal .	Maisy, Yuze, Yujun
Reward construction	Unified Affective Matrix Reward Engine integration	Integrate the Unified Affective Matrix Reward Engine directly into the GRPOTrainer to provide real-time clinical alignment and safety rewards .	Maisy, Yuze, Yujun
Reward construction	Verifier (reward scorer)	Options: -Deploy PsychBERT and/or sentiment-classification-bert-mini to act as a reward scorer ensuring "Chosen" responses are scored appropriately within the Unified Affective Matrix Reward Engine -Use of Weaver (ensemble methods)	Maisy, Yuze, Yujun
Model training	QLoRA Training (setup)	Configure the environment using Unsloth and TRL to enable QLoRA training on a single 24GB GPU	Yujun, Yifan, Akshaya
Model training	DPO Pipeline Implementation (optional if we can skip to GRPO)	Implement the DPO pipeline to use the data generated by the simulator + the reward signals generated by the verifier on simulator-generated data	Yujun, Yifan, Akshaya
Model training	GRPO Pipeline Implementation	Implement the GRPO pipeline to use the data generated by the simulator + the reward signals generated by the verifier on simulator-generated data	Yujun, Yifan, Akshaya
Model training	MentalChat16K data processing	Analyze and preprocess the MentalChat16K dataset, developing automatic ways to evaluate response quality including empathy, emotional tone, and safety, and create preference-style datasets that can be used for later DPO-style training.	Yujun, Yifan, Akshaya
Model evaluation	Validation of reward construction signal	Validation that training model using the Unified Matrix Reward Engine yield clinically significant results Options: -Clinical Calibration: Apply Isotonic Regression to calibrate automated scores against human-expert ratings, ensuring the AI's judgment matches a real therapist.	Chenchen, Wenxi
Model evaluation	Safety of model outputs	A few ideas: -Domain-Specific Constitutional AI: Enhancing Safety in LLM-Powered Mental Health Chatbots: interesting use of RLAIF with constitutional AI to improve safety -https://github.com/w-is-h/psychosis-bench: this benchmark can be used to validate the effectiveness of RL in reducing the psychosis risk -https://mental.jmir.org/2025/1/e75078/ -https://www.sciencedirect.com/science/article/abs/pii/S0920996425002816 -https://dl.acm.org/doi/10.1145/3715275.3732039 -https://pmc.ncbi.nlm.nih.gov/articles/PMC12550315/#bb0045 (see section 6)	Chenchen, Wenxi