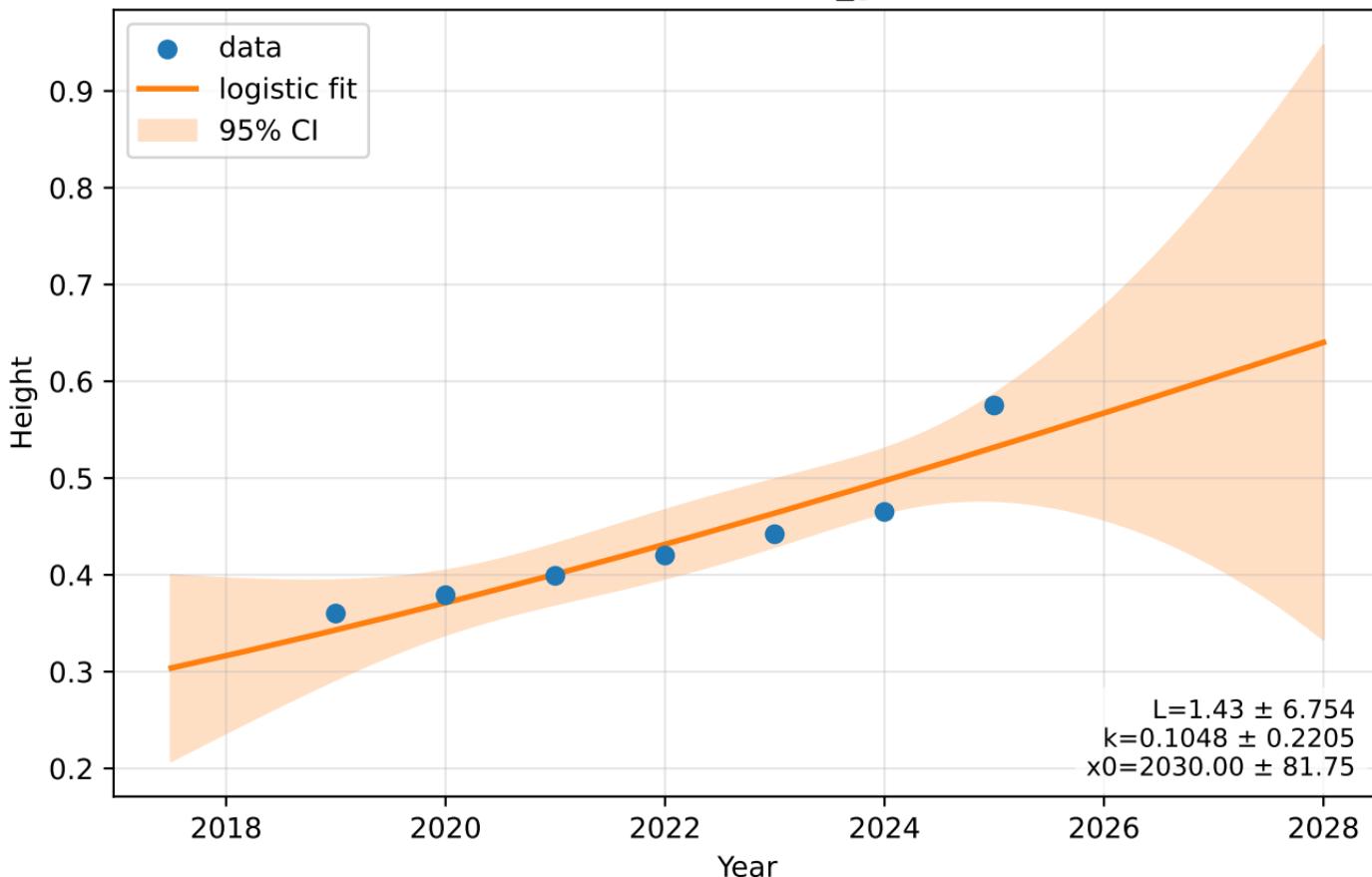
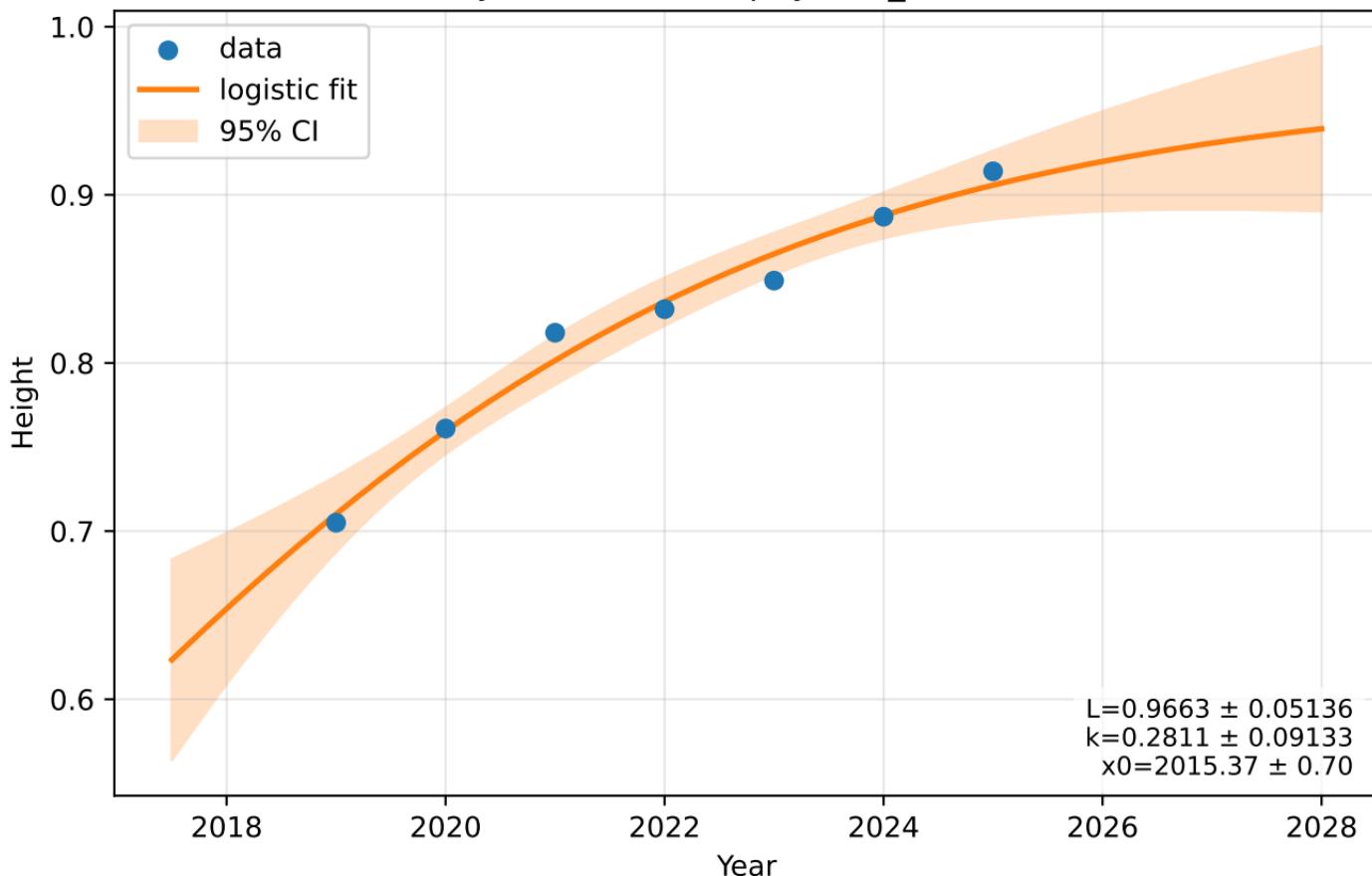


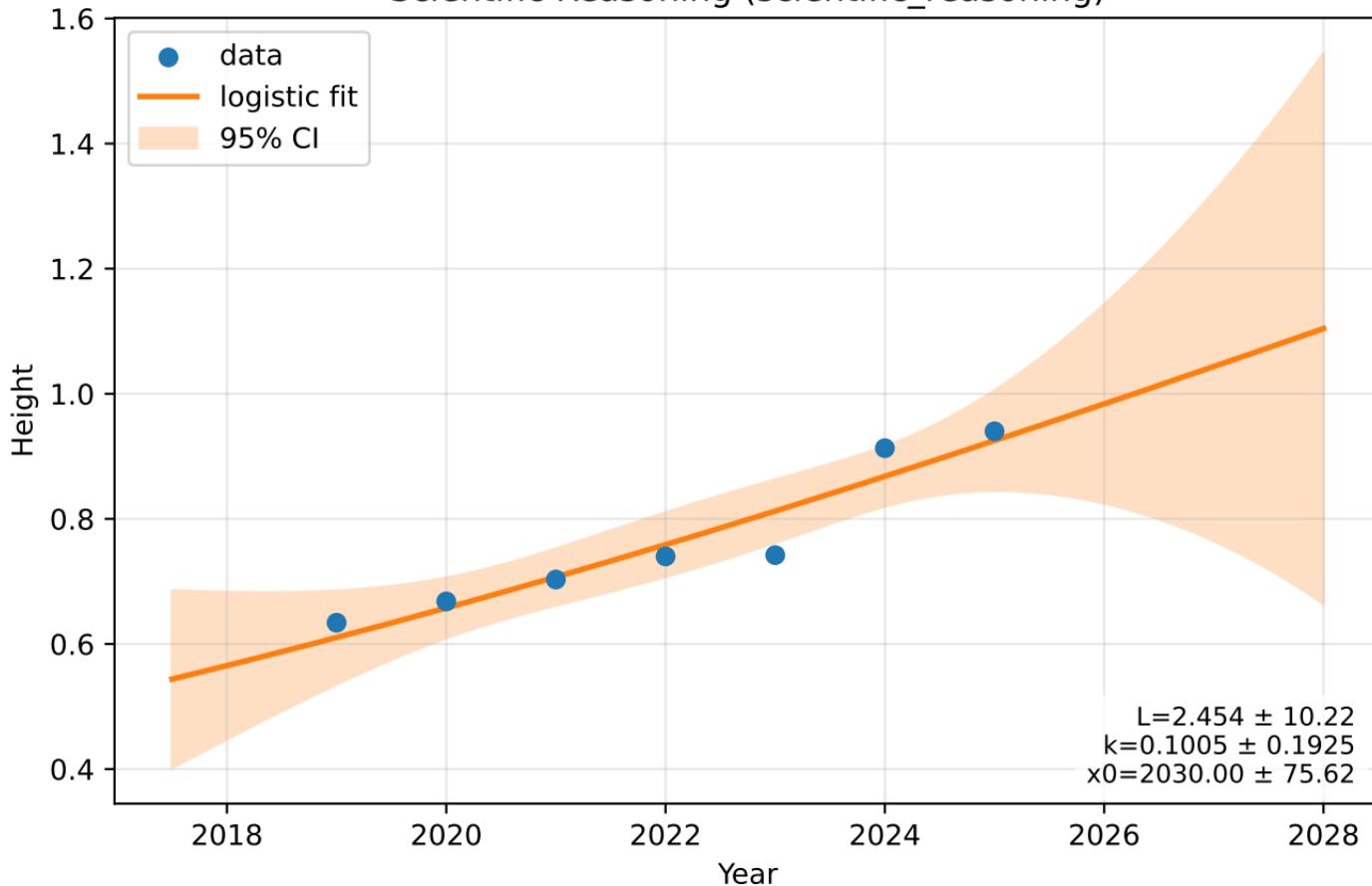
Code Generation (code_generation)



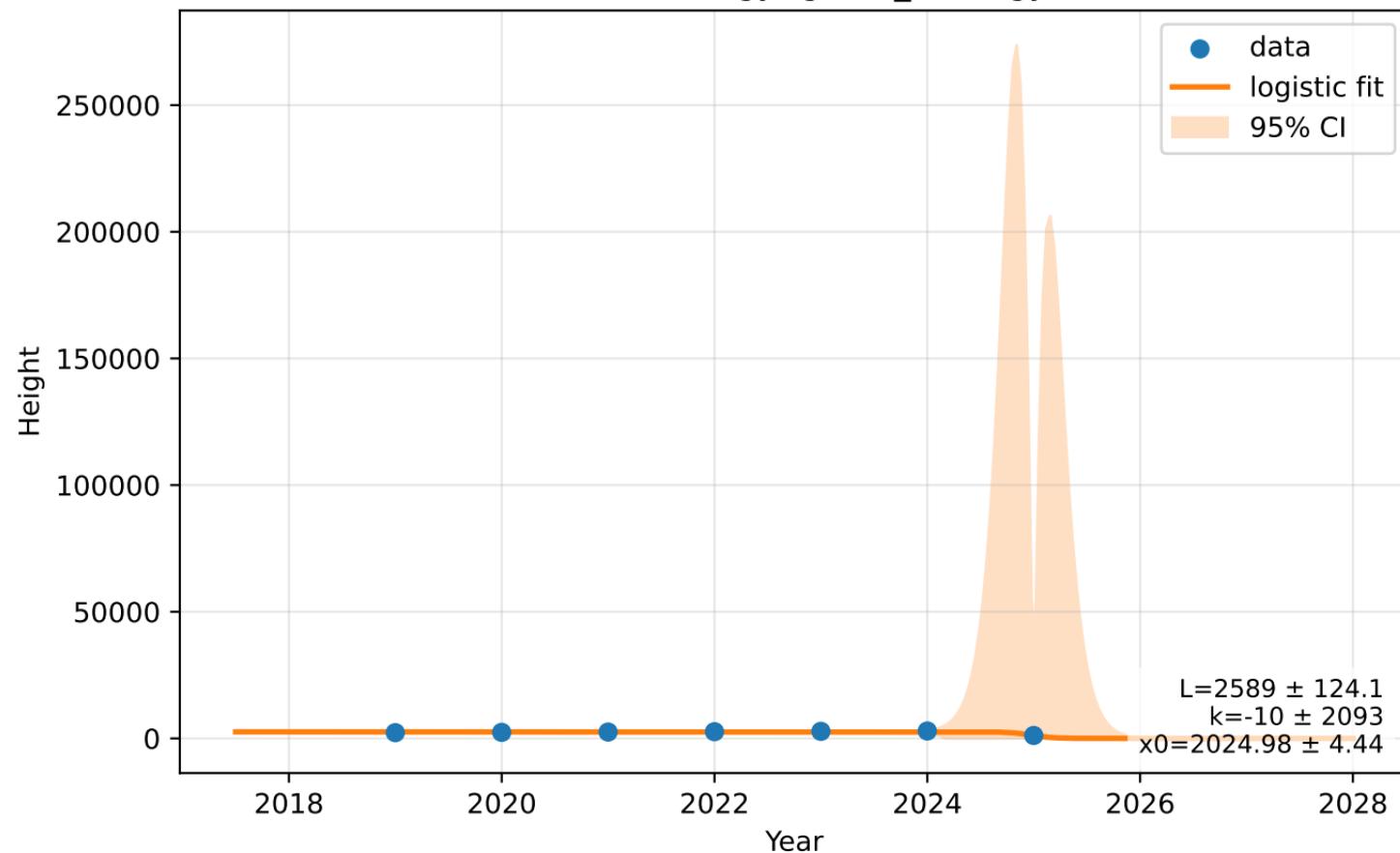
Physical Intuition (physical_intuition)



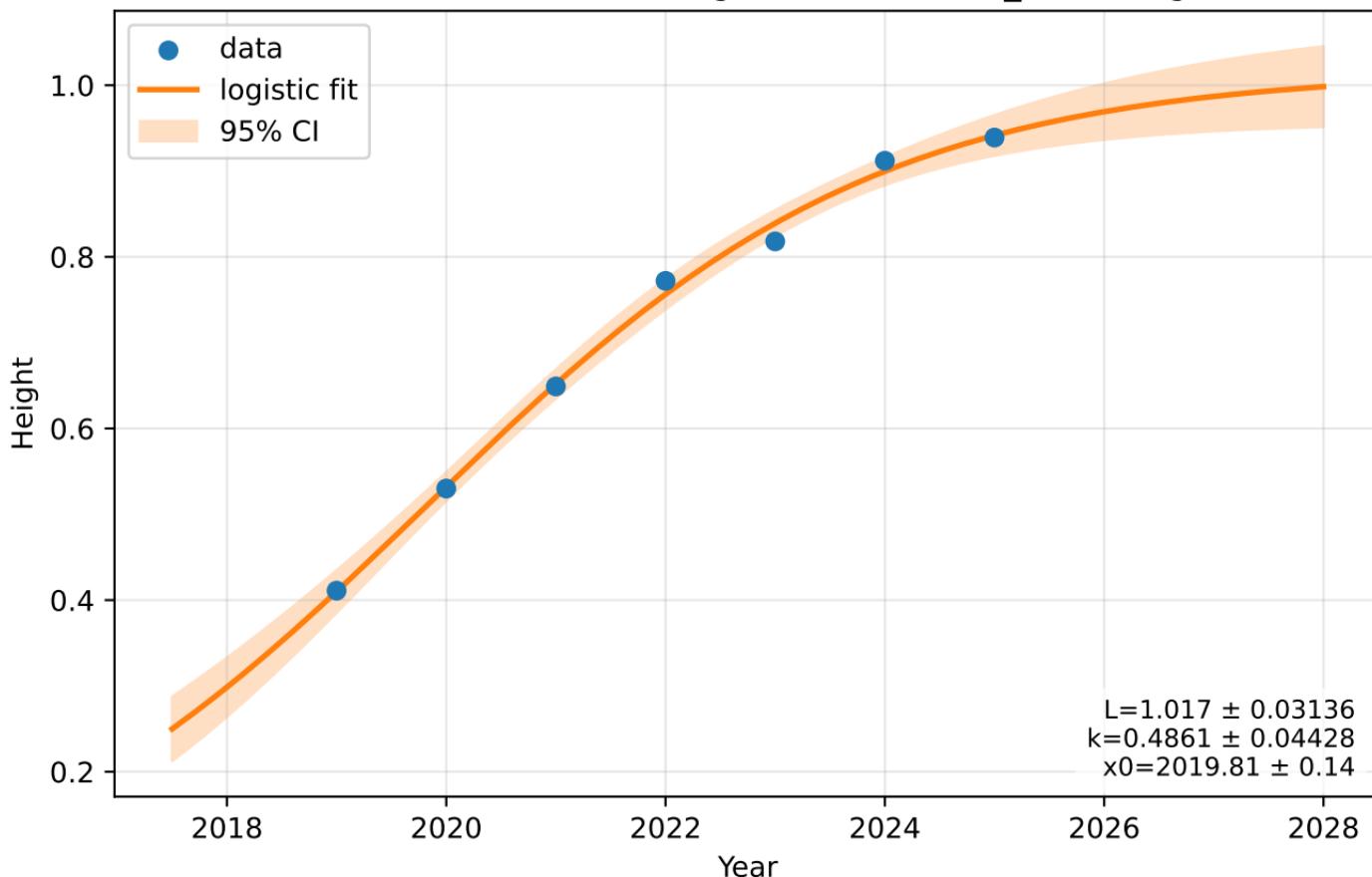
Scientific Reasoning (scientific_reasoning)



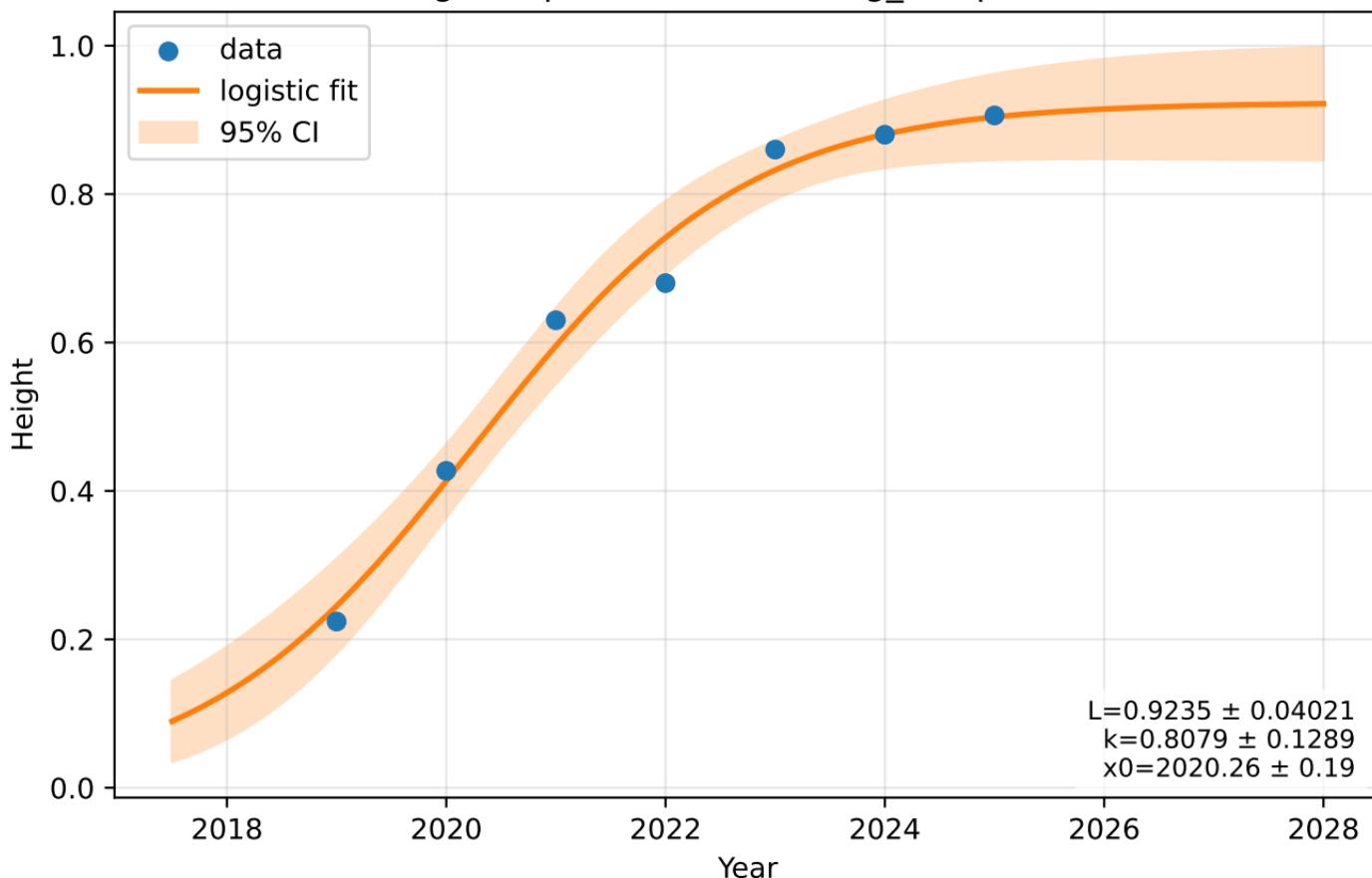
Game Strategy (game_strategy)



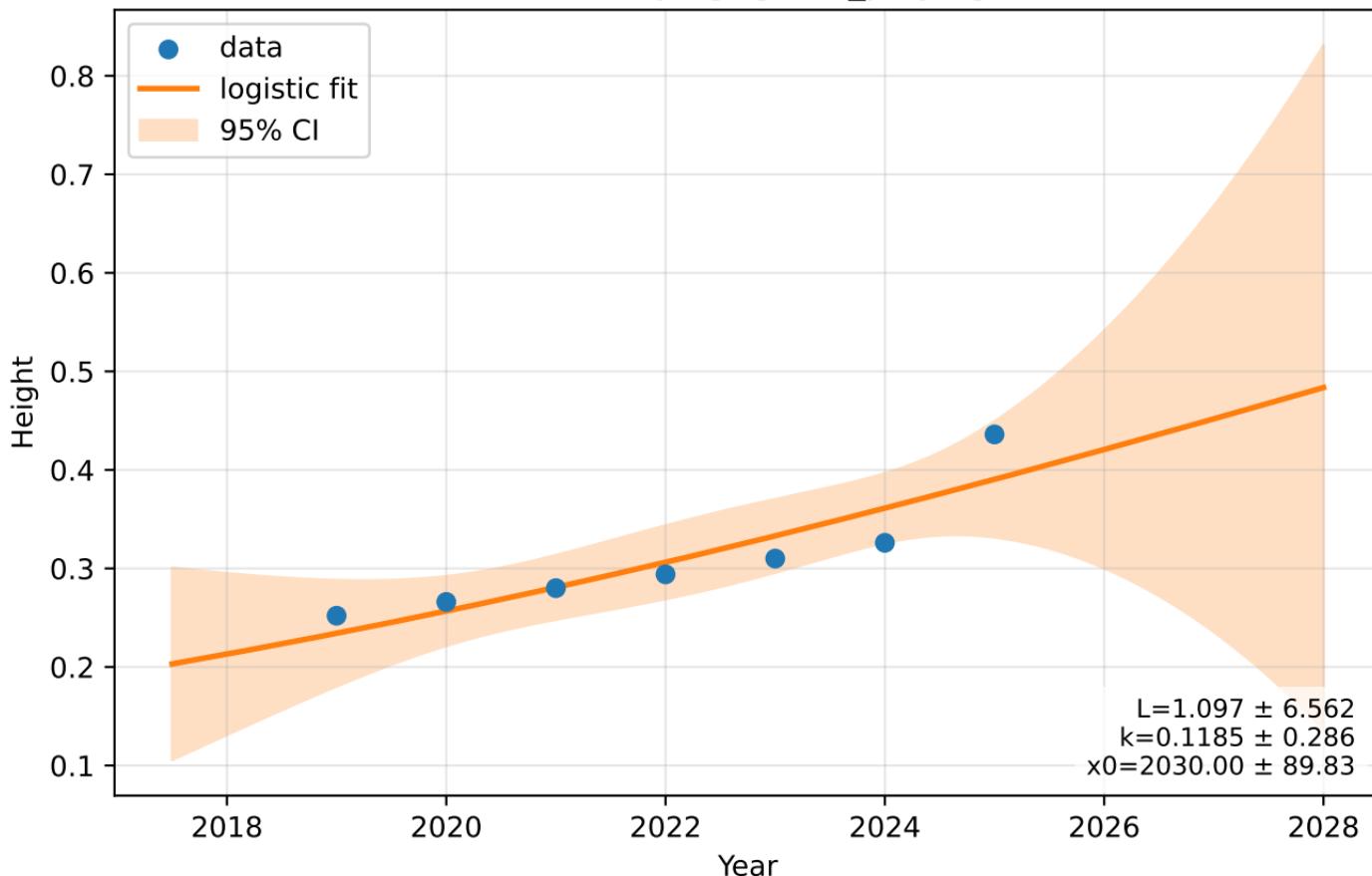
Commonsense Reasoning (commonsense_reasoning)



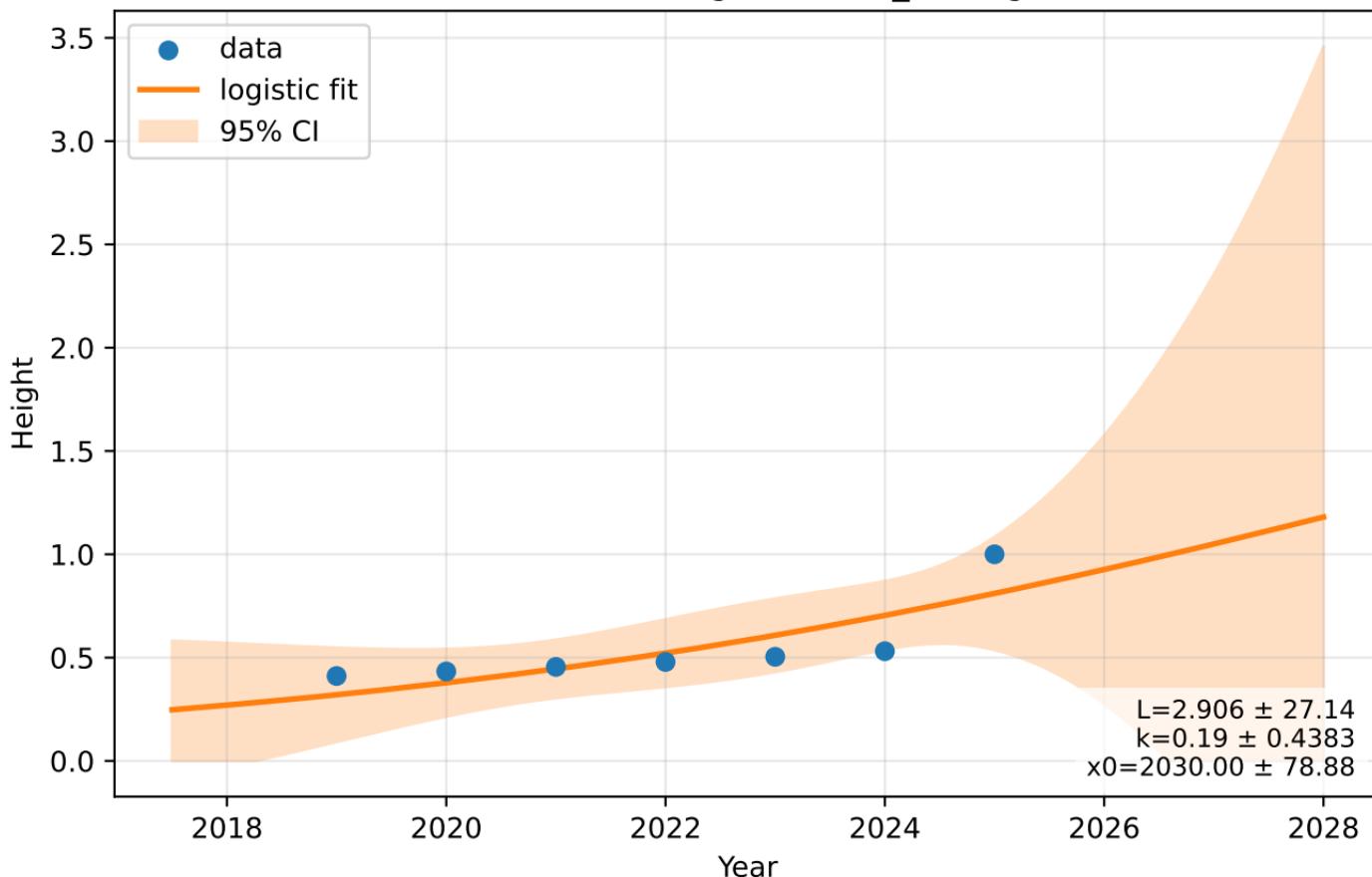
Reading Comprehension (reading_comprehension)



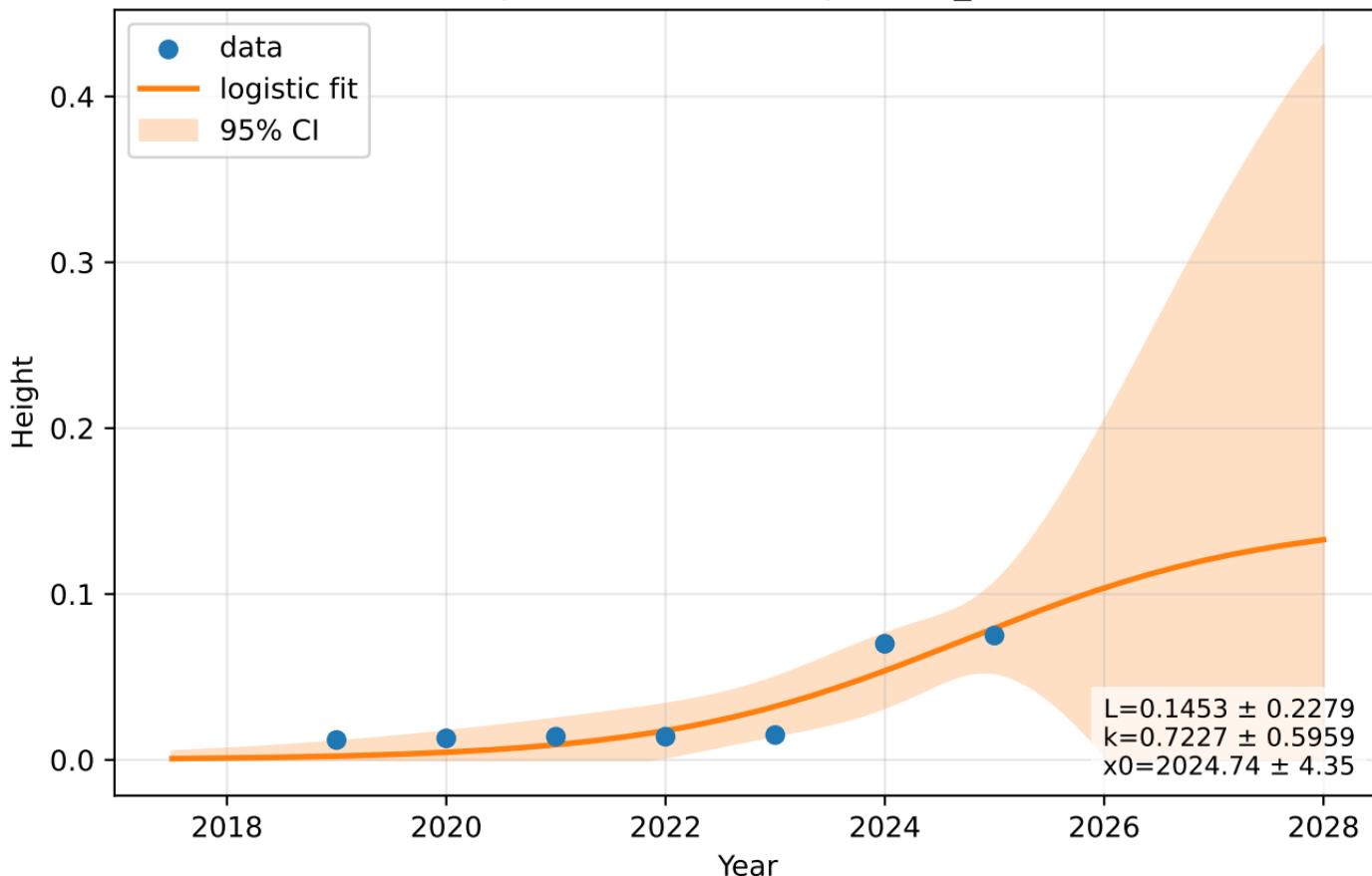
Game Playing (game_playing)



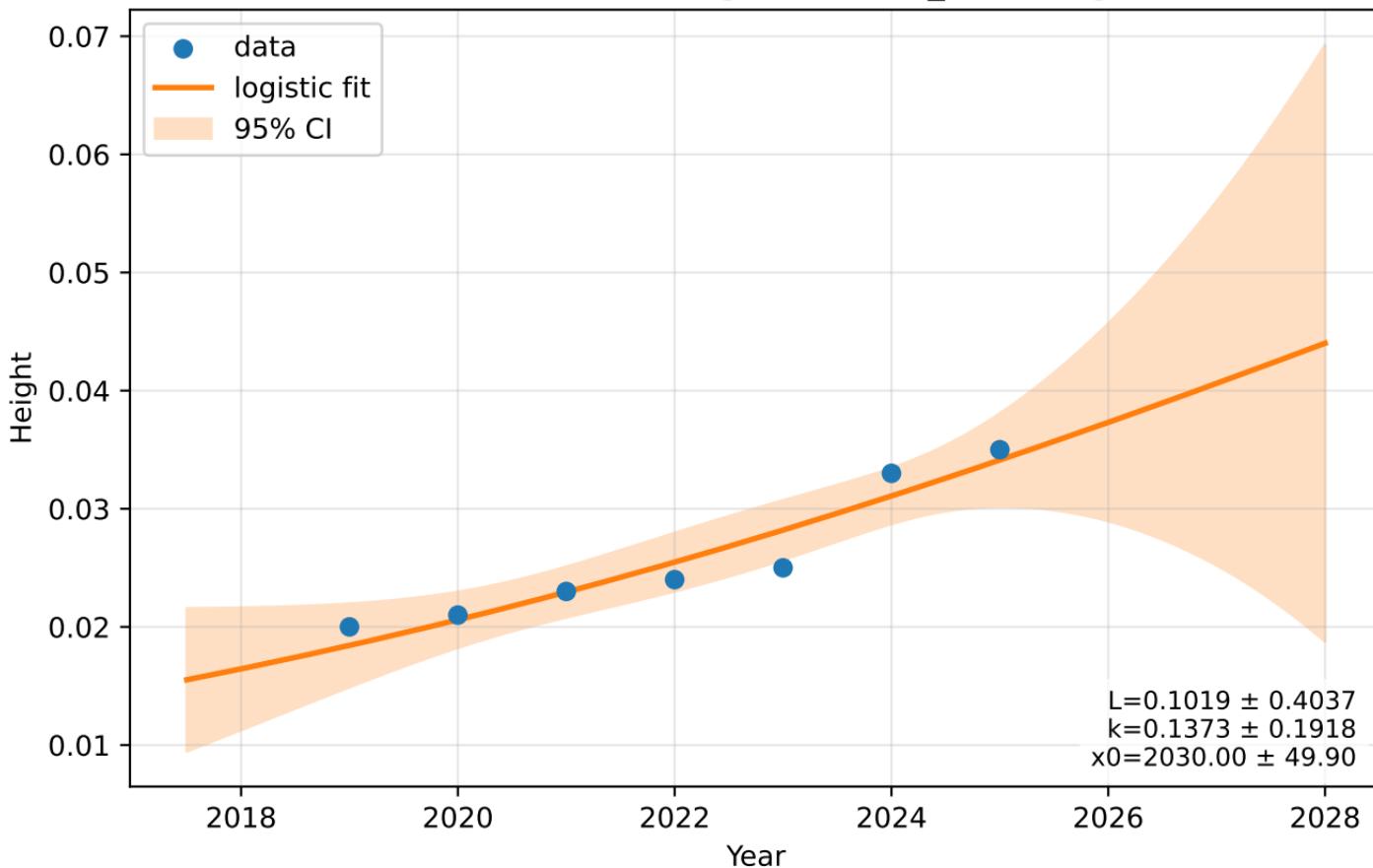
Creative Writing (creative_writing)



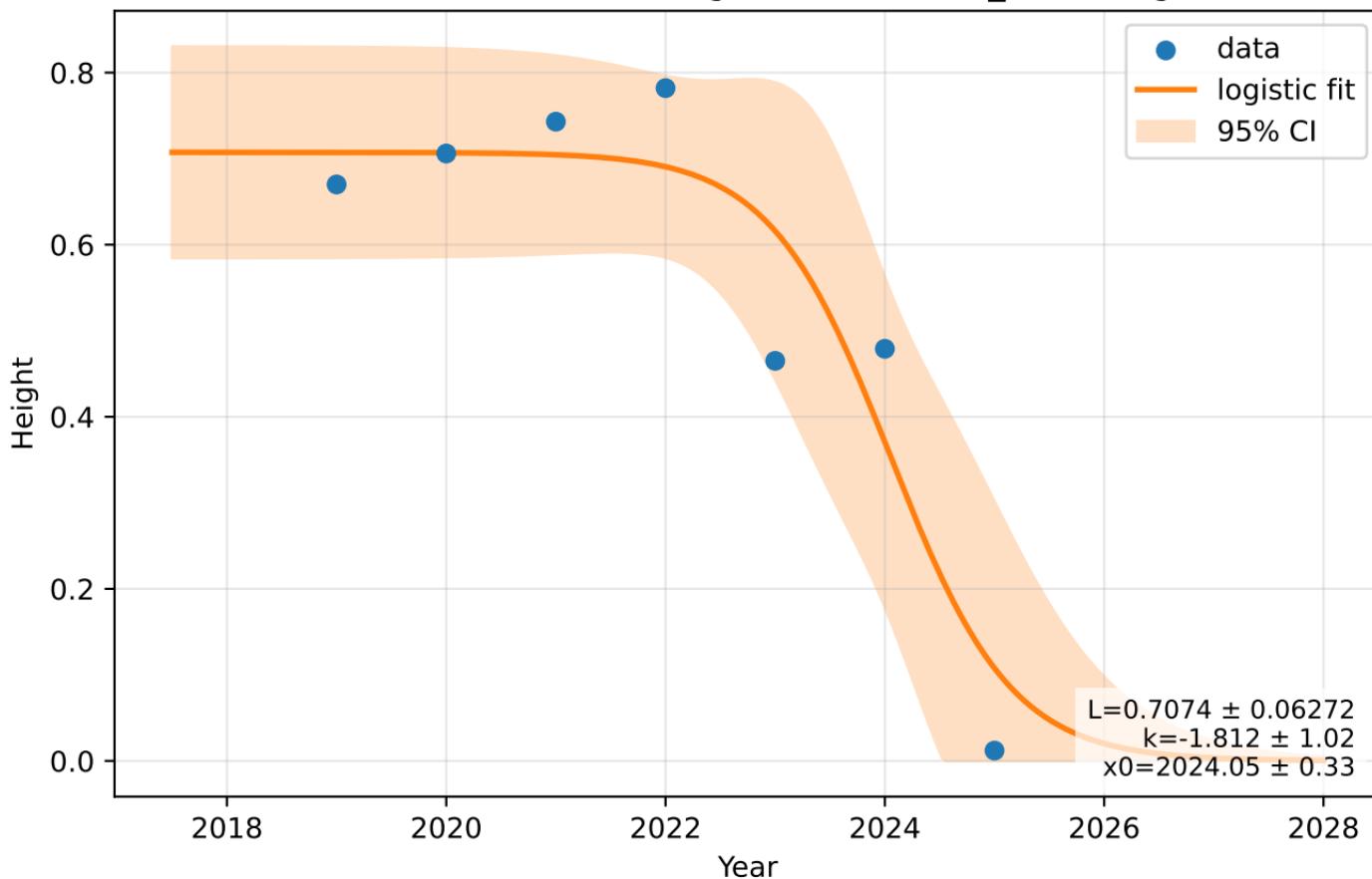
Competition Math (competition_math)



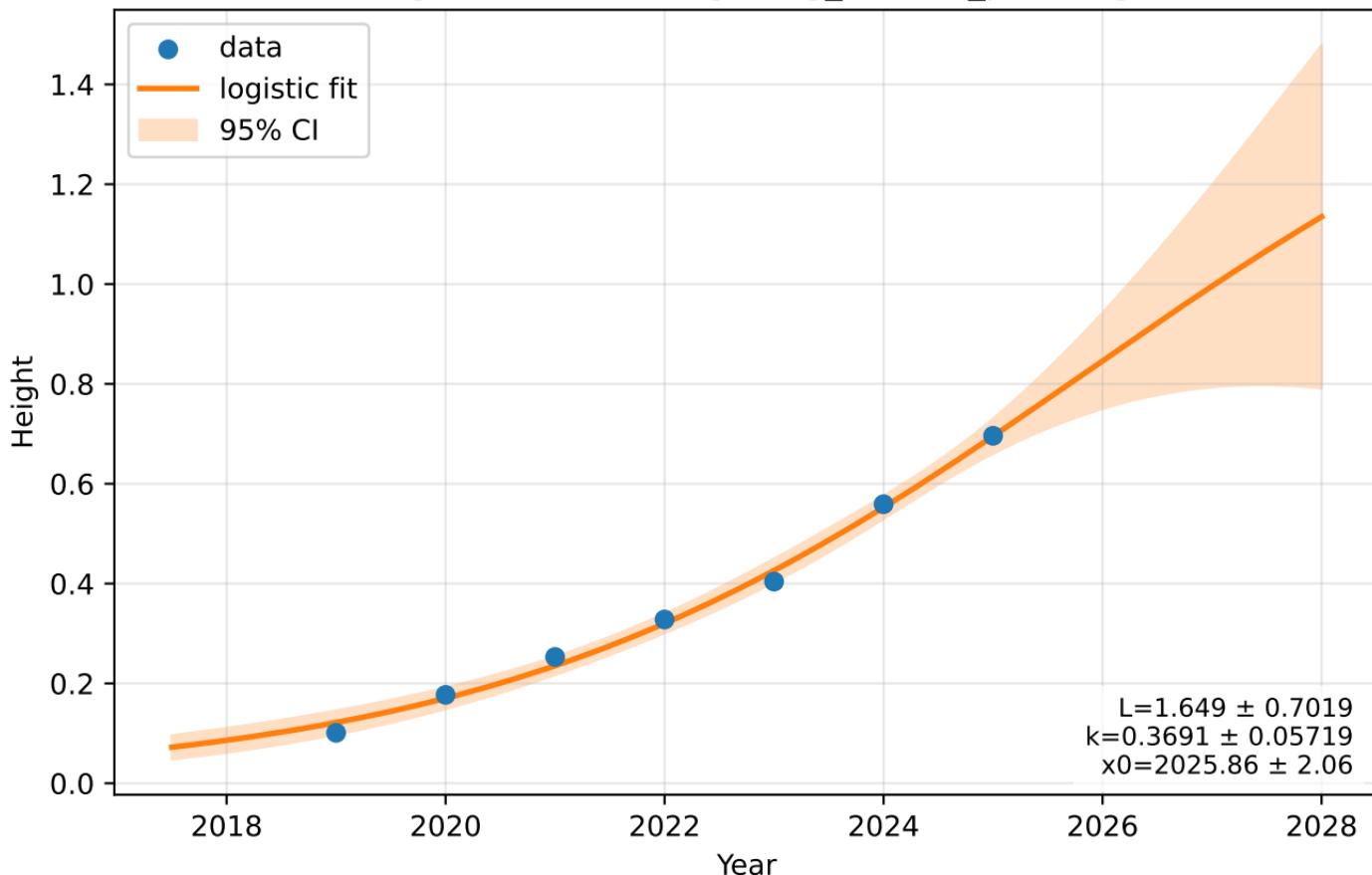
Advanced Reasoning (advanced_reasoning)



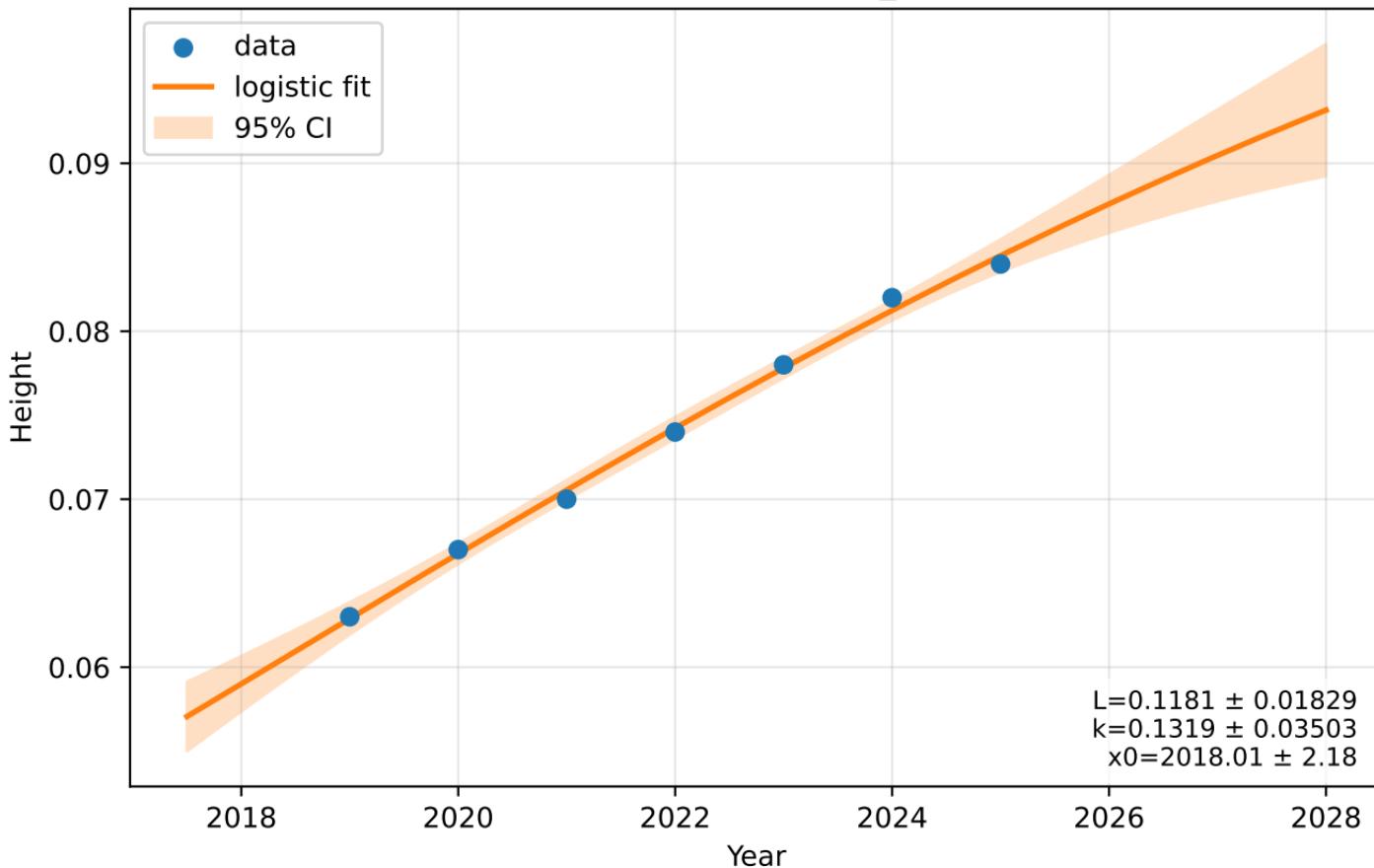
Mathematical Reasoning (mathematical_reasoning)



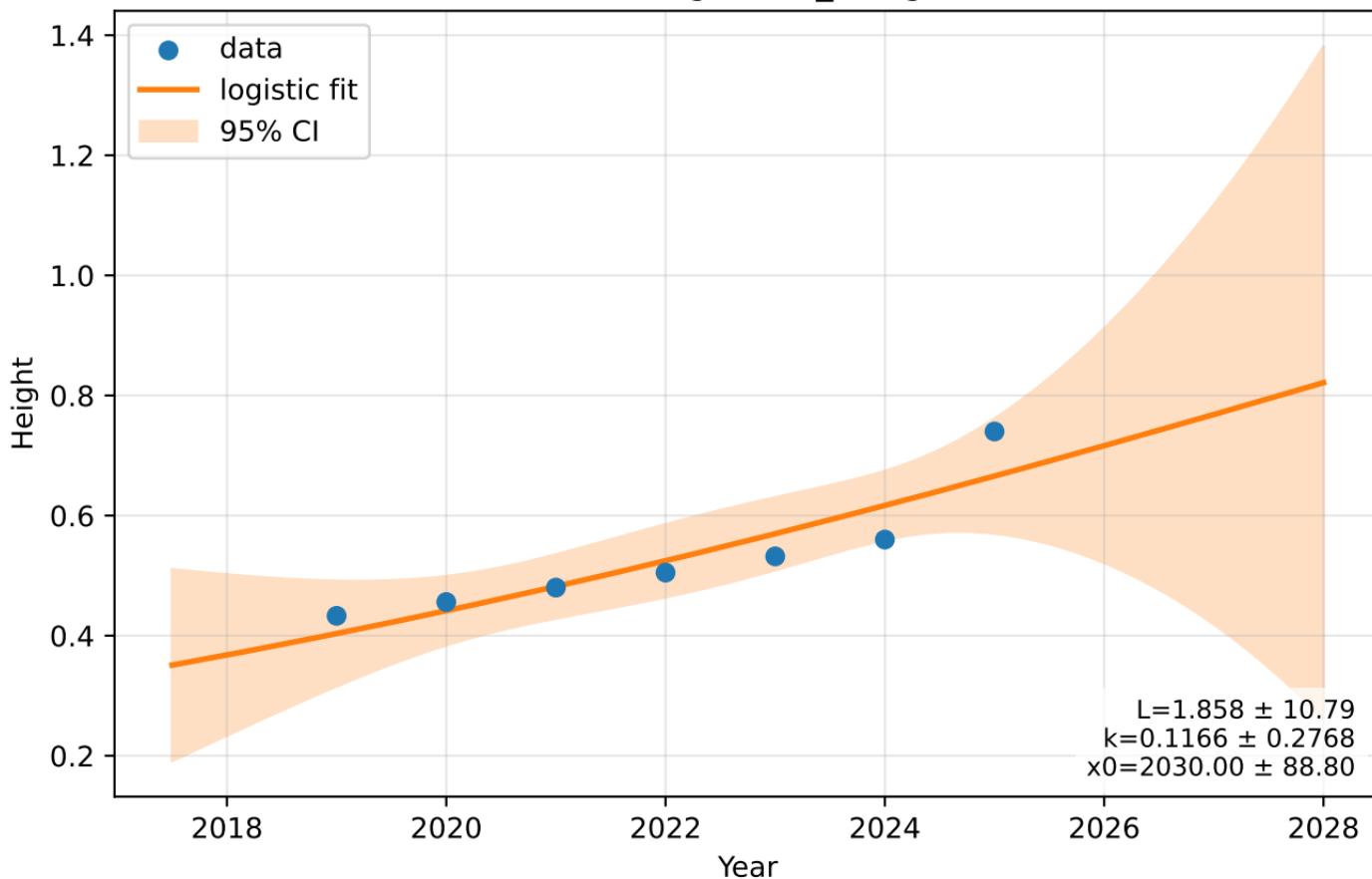
Long Horizon Planning (long_horizon_planning)



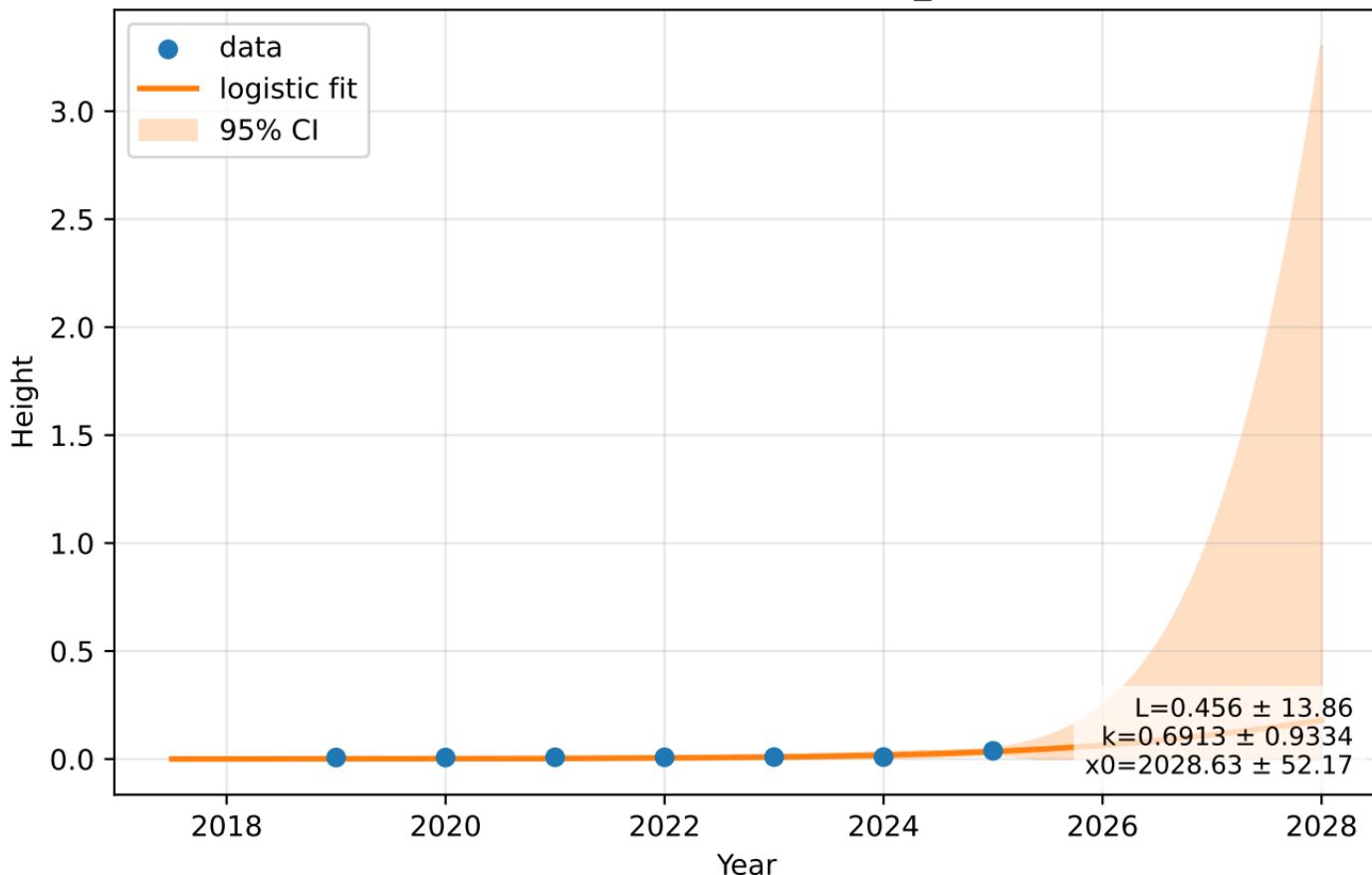
Writing Quality (writing_quality)



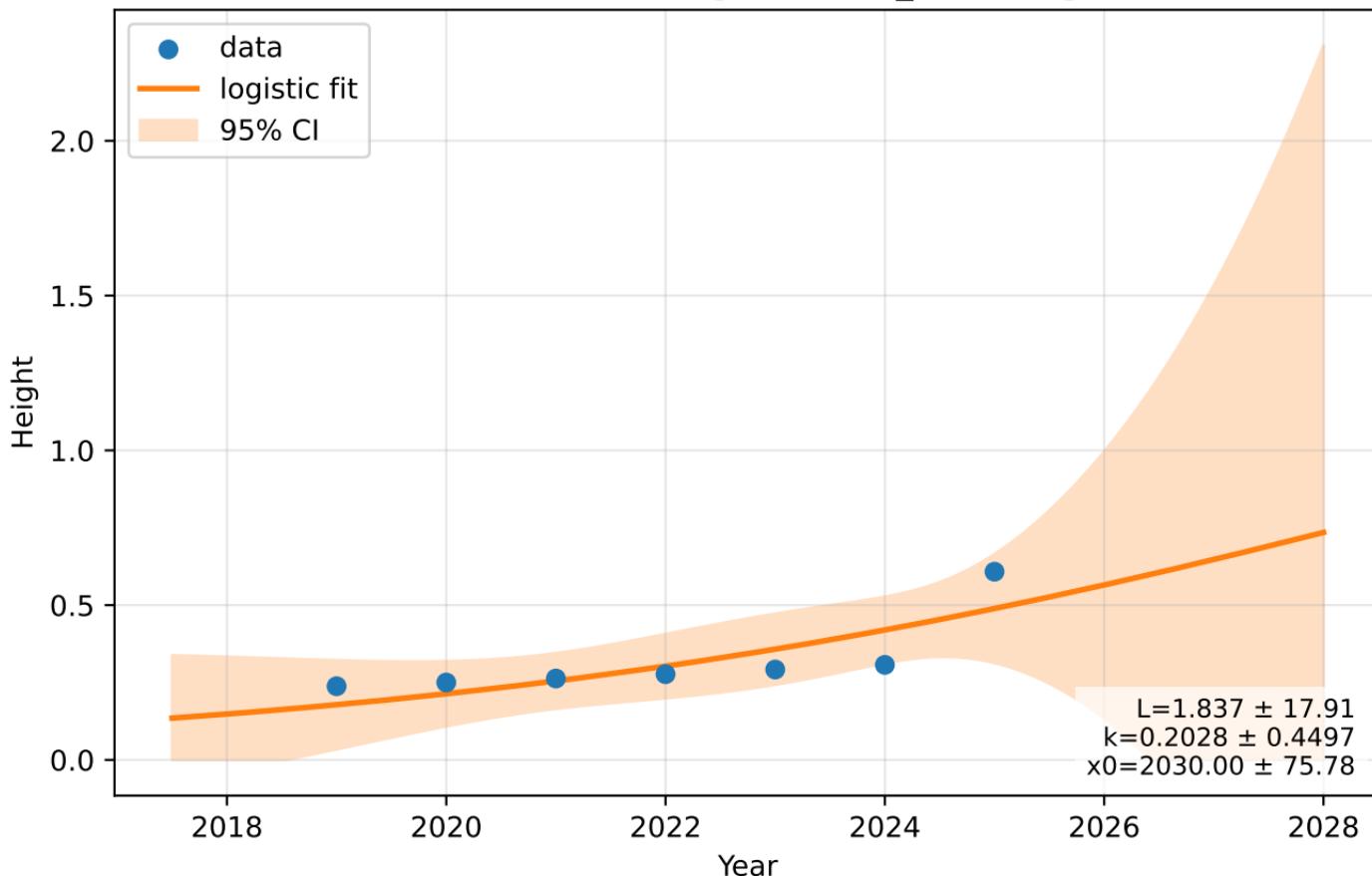
Cad Design (cad_design)



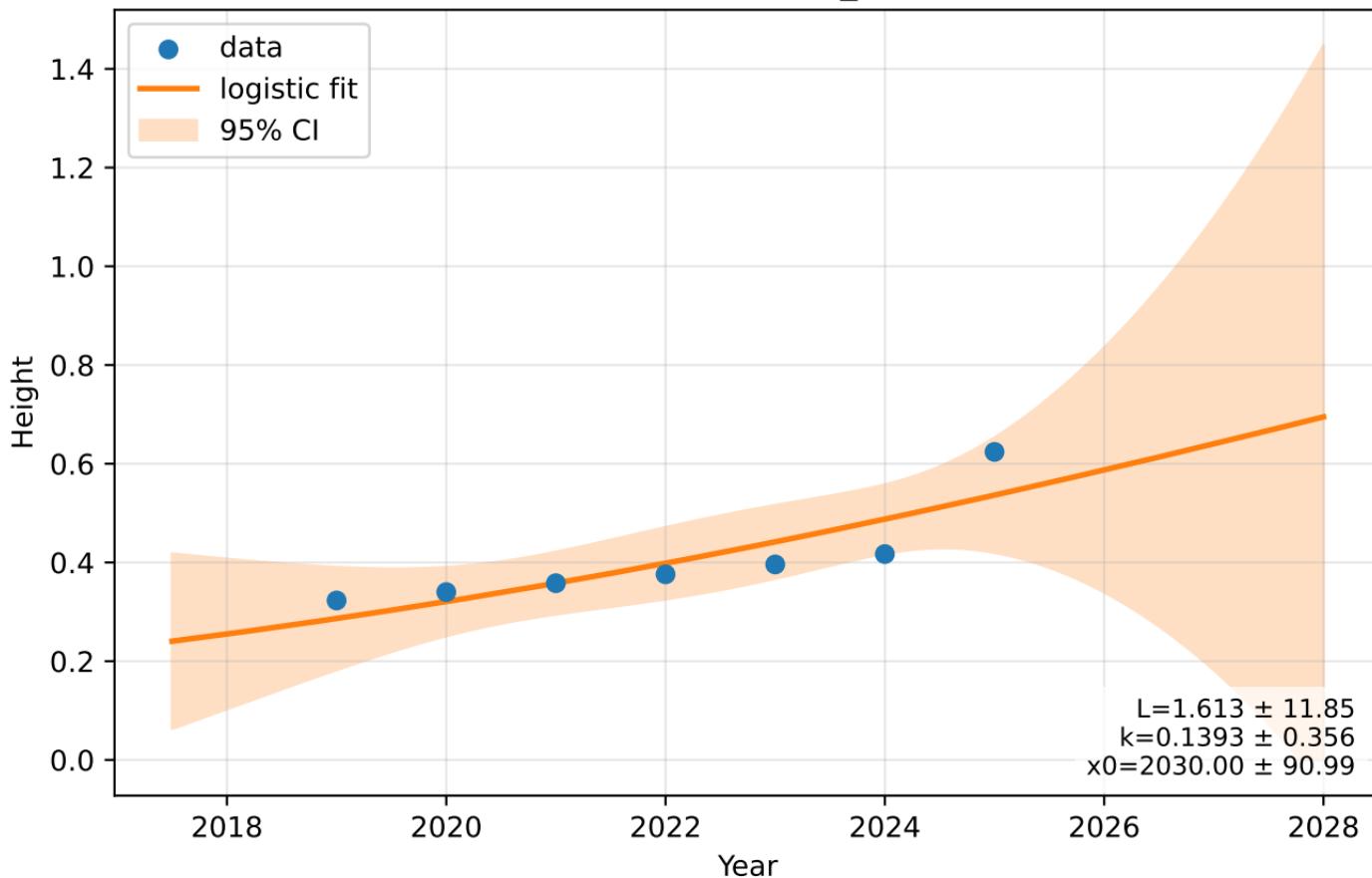
Advanced Mathematics (advanced_mathematics)



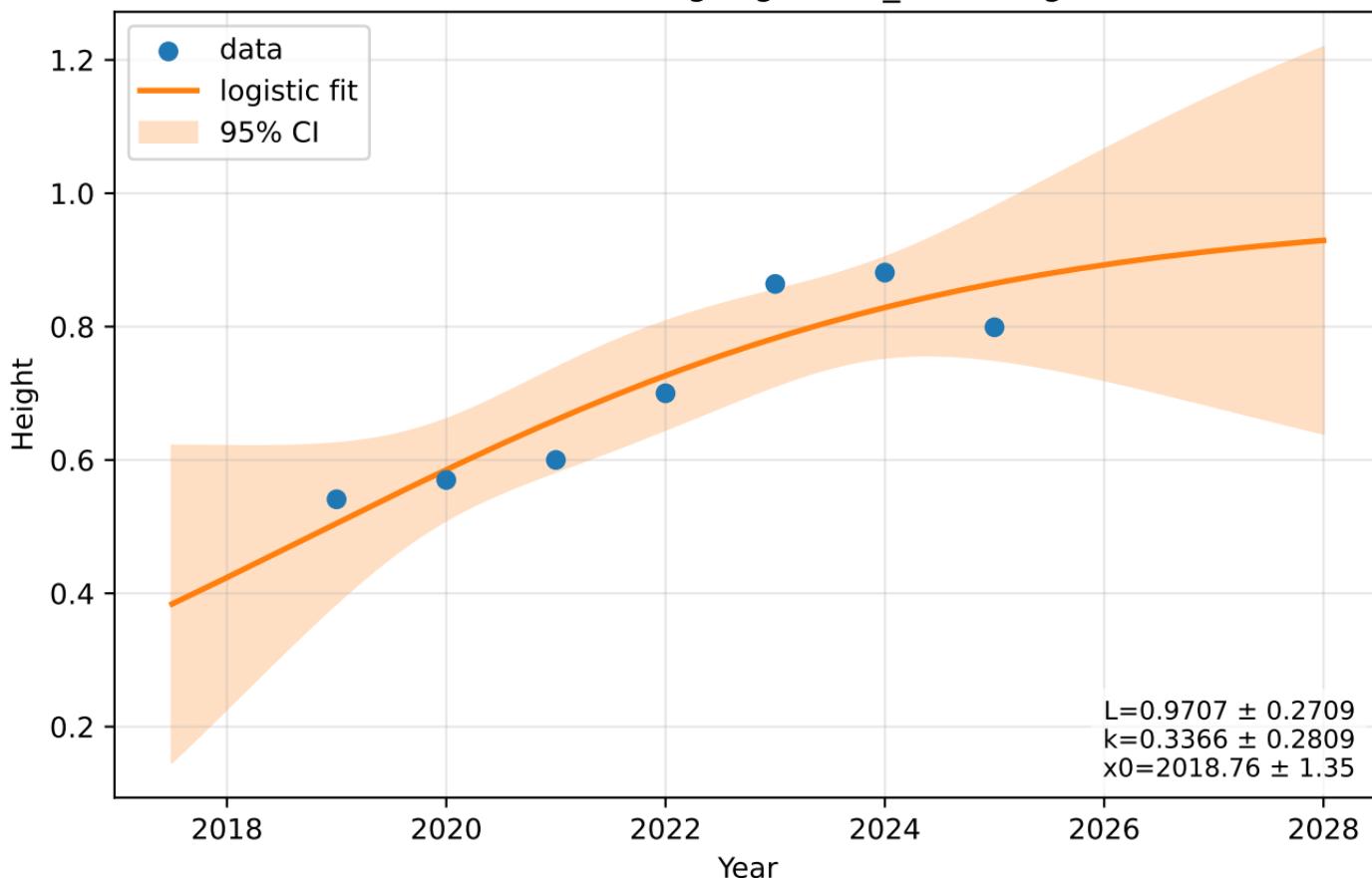
Abstract Reasoning (abstract_reasoning)



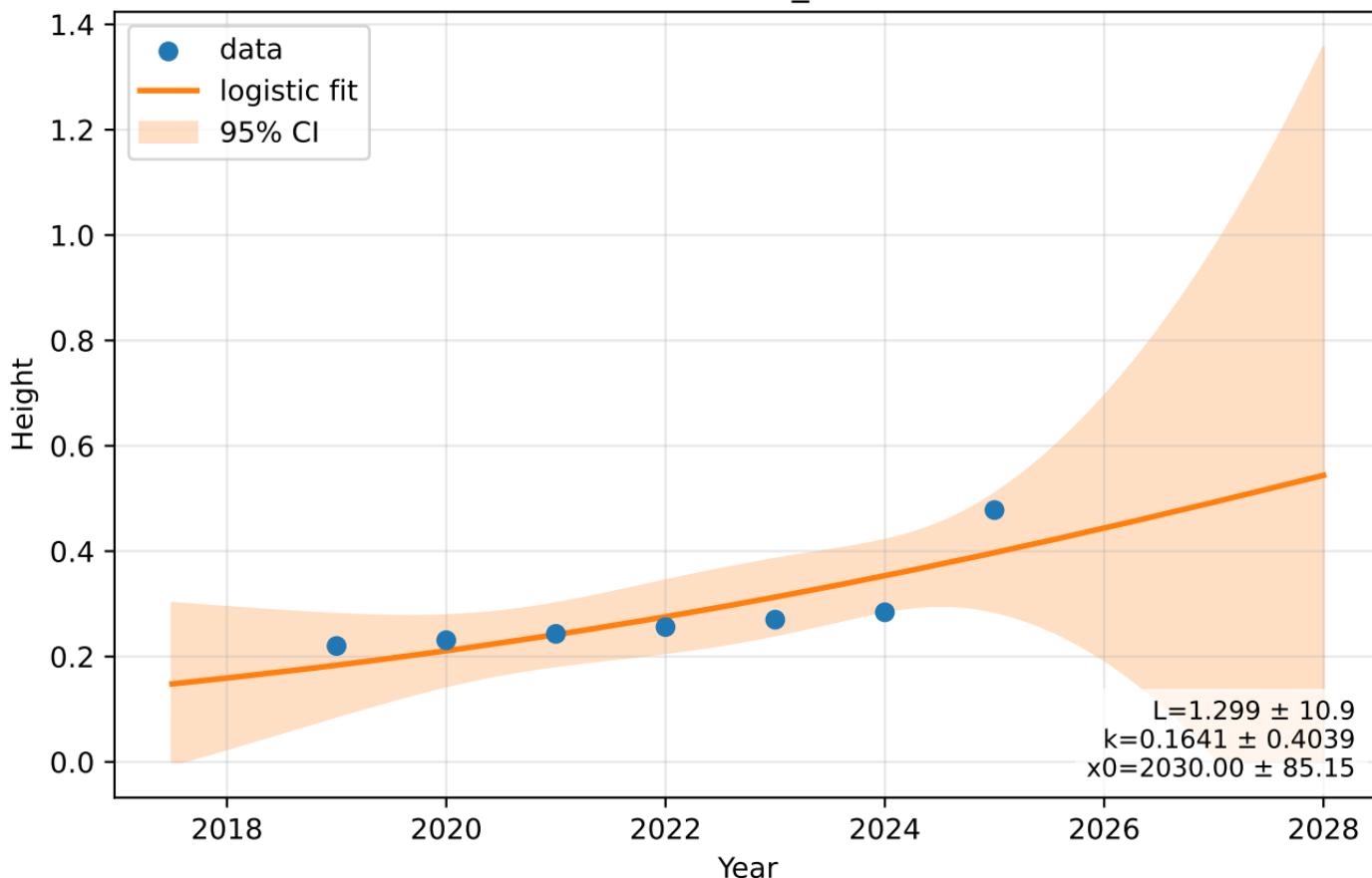
Basic Tasks (basic_tasks)



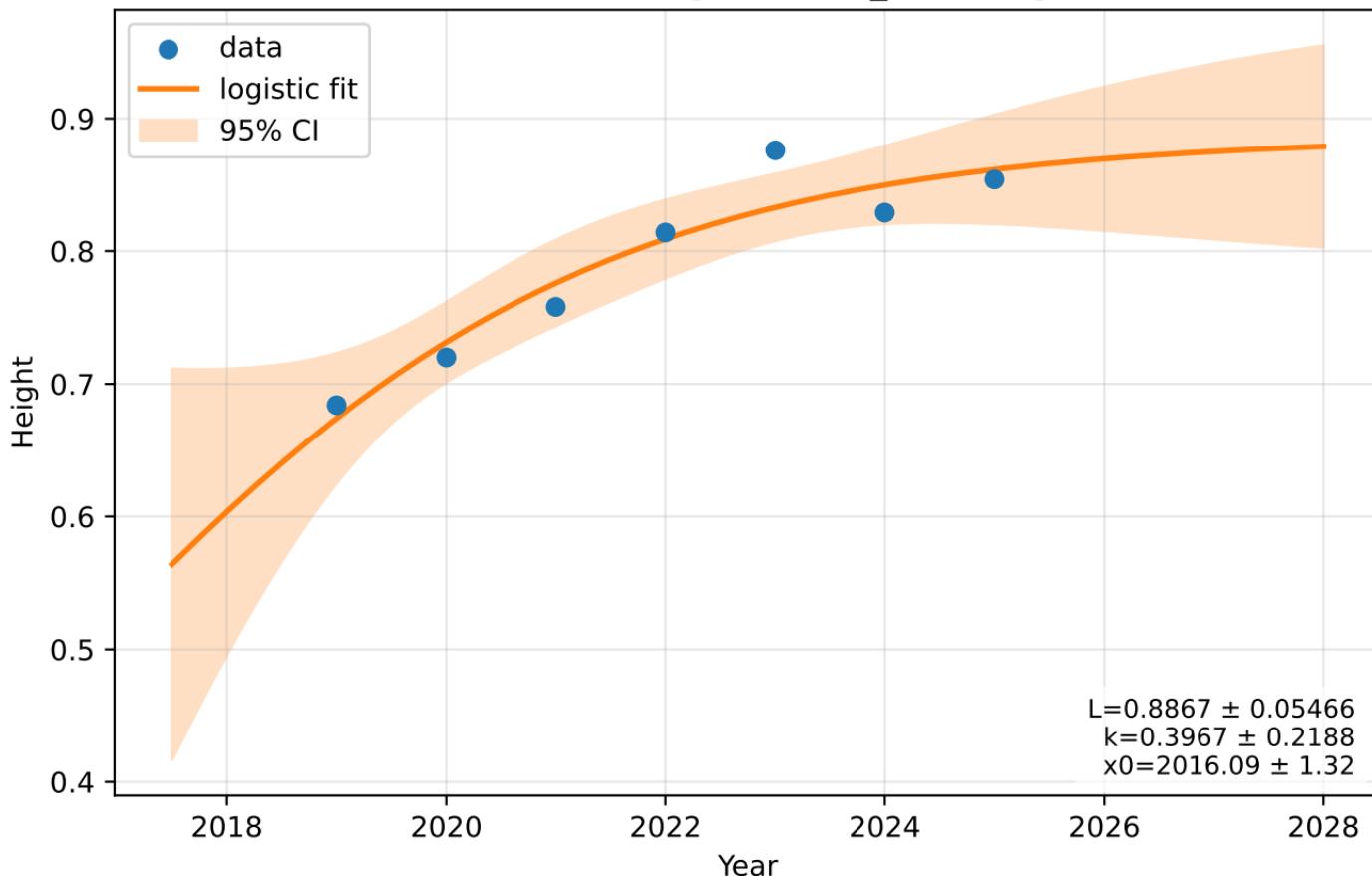
General Knowledge (general_knowledge)



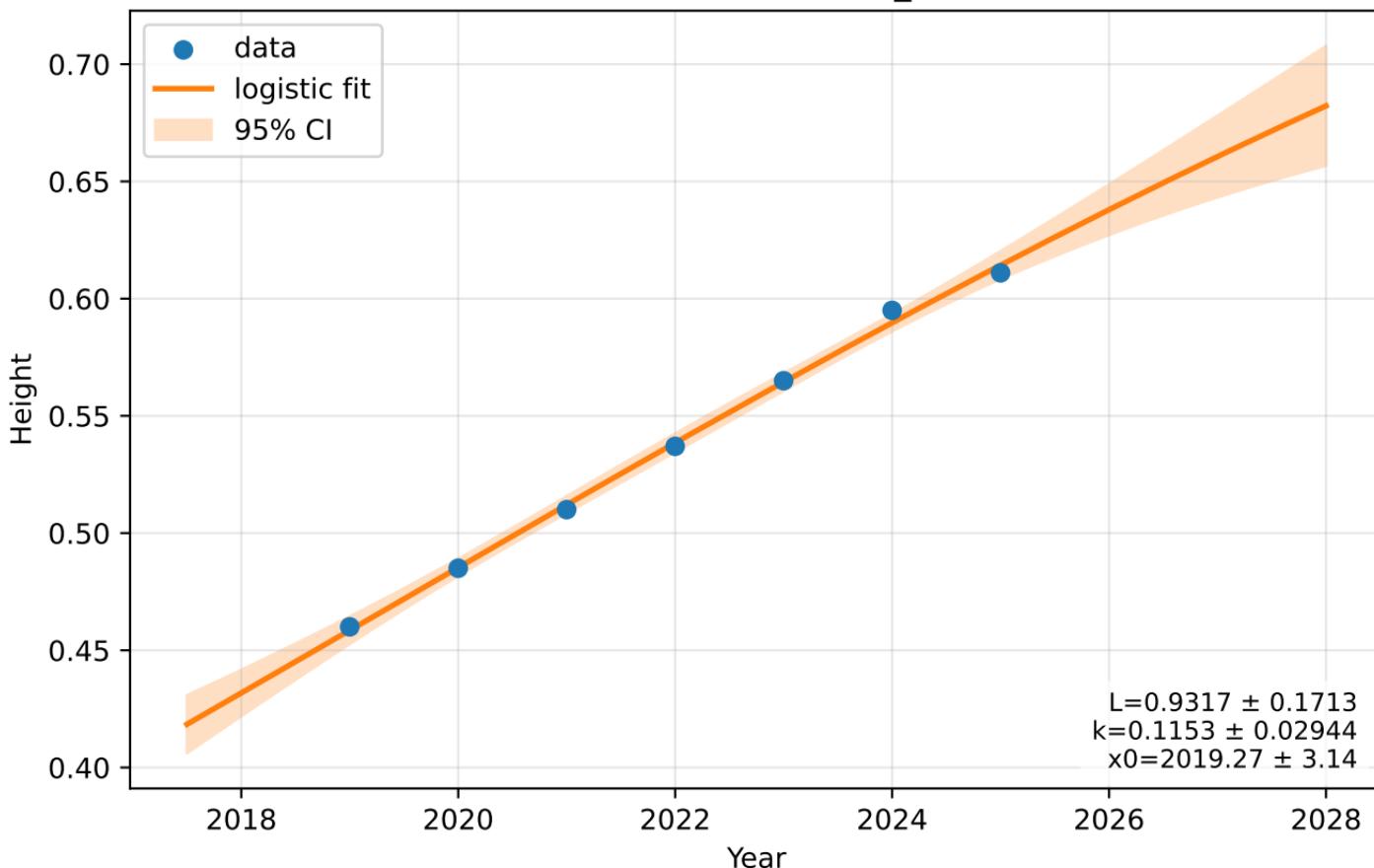
Os Interaction (os_interaction)



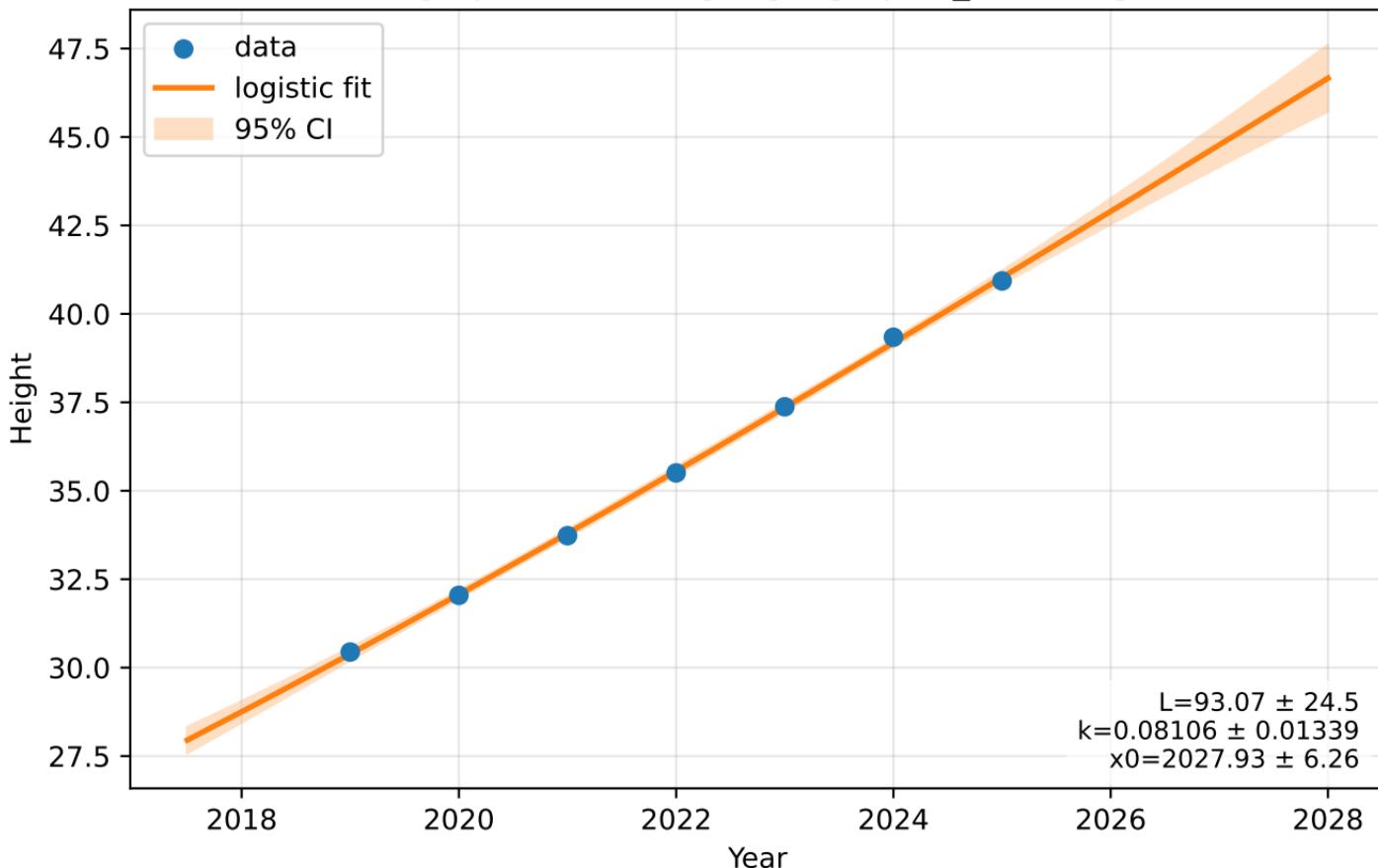
Factual Knowledge (factual_knowledge)



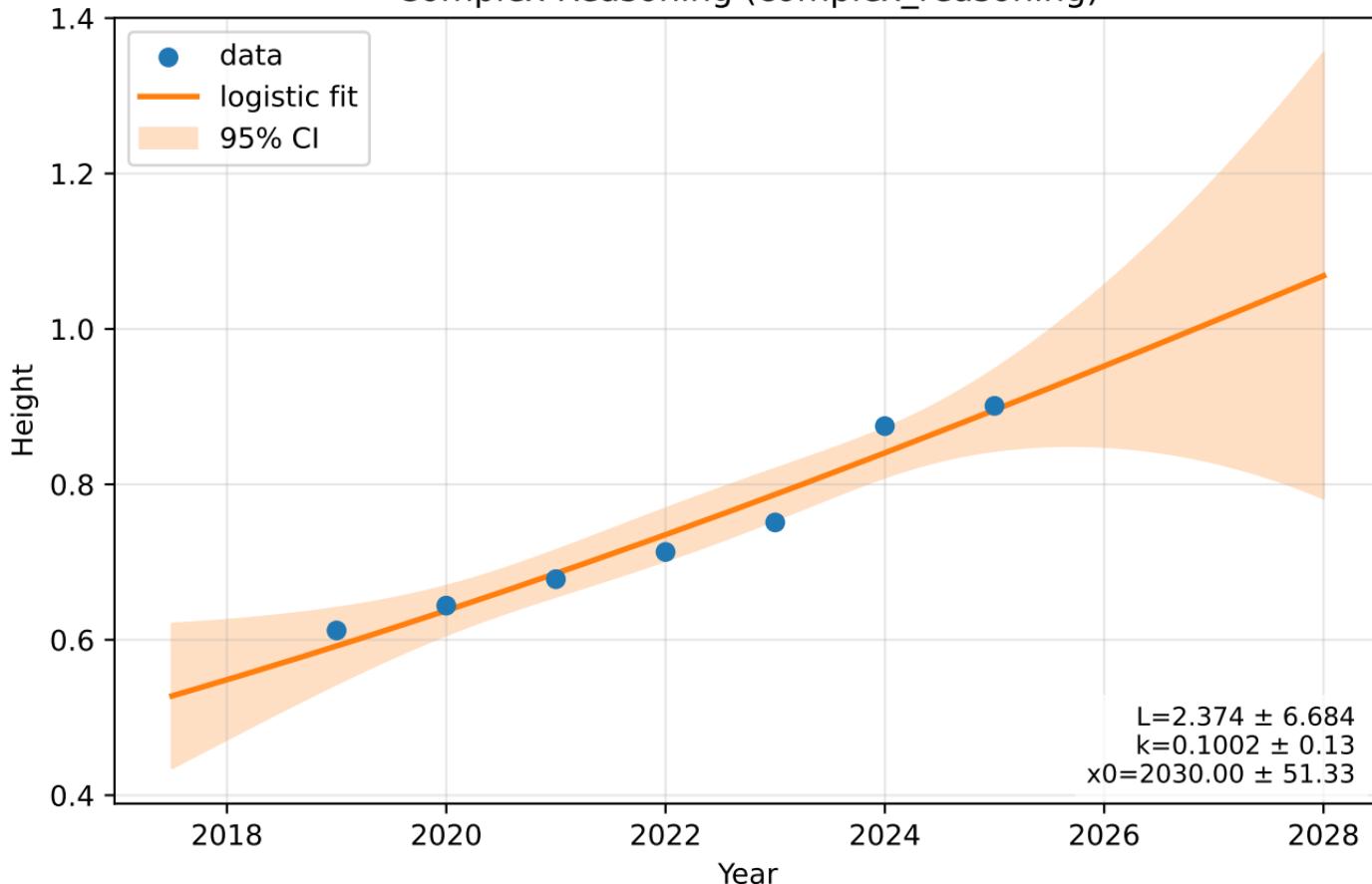
Unusual Tasks (unusual_tasks)



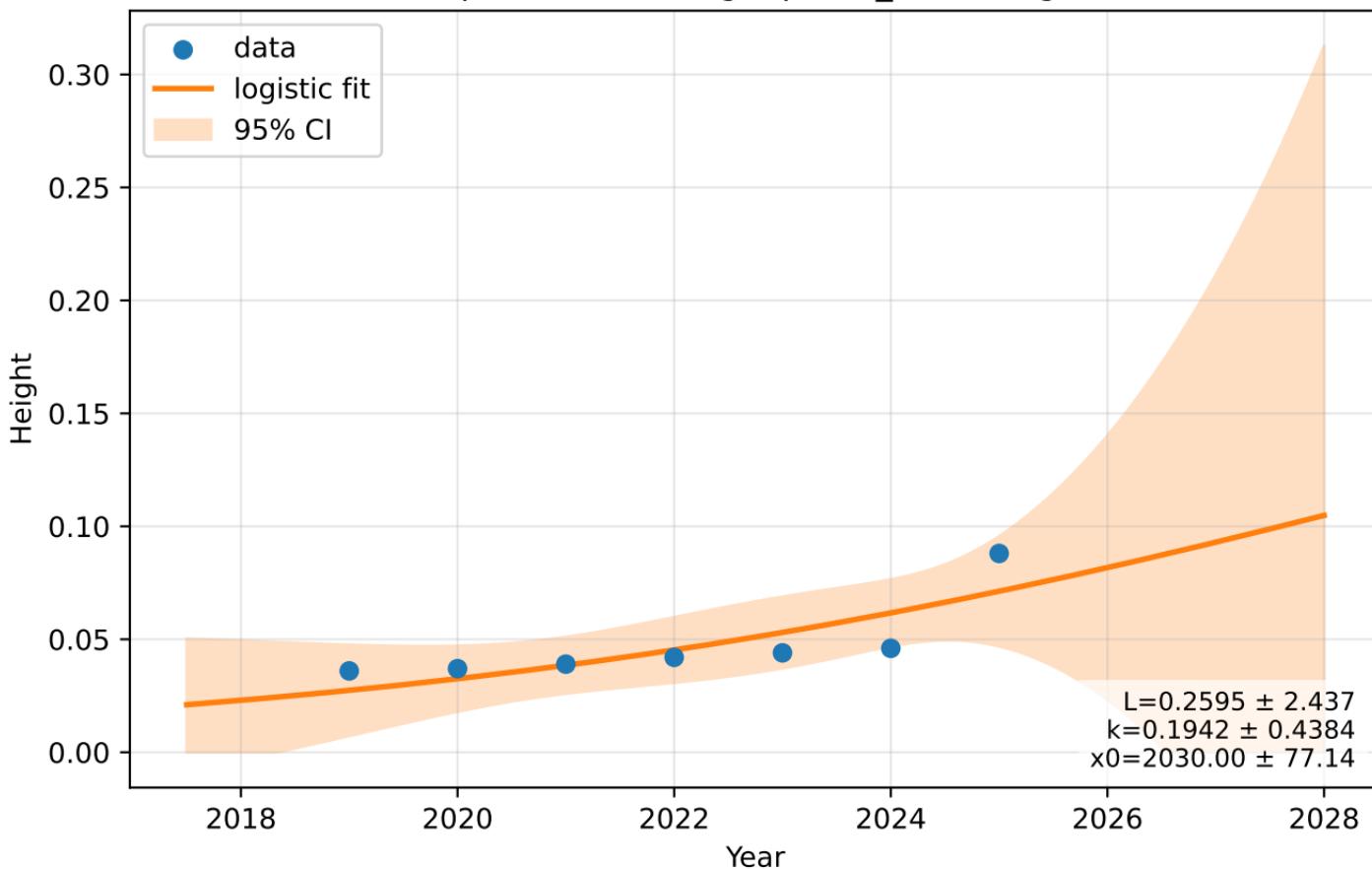
Geographic Knowledge (geographic_knowledge)



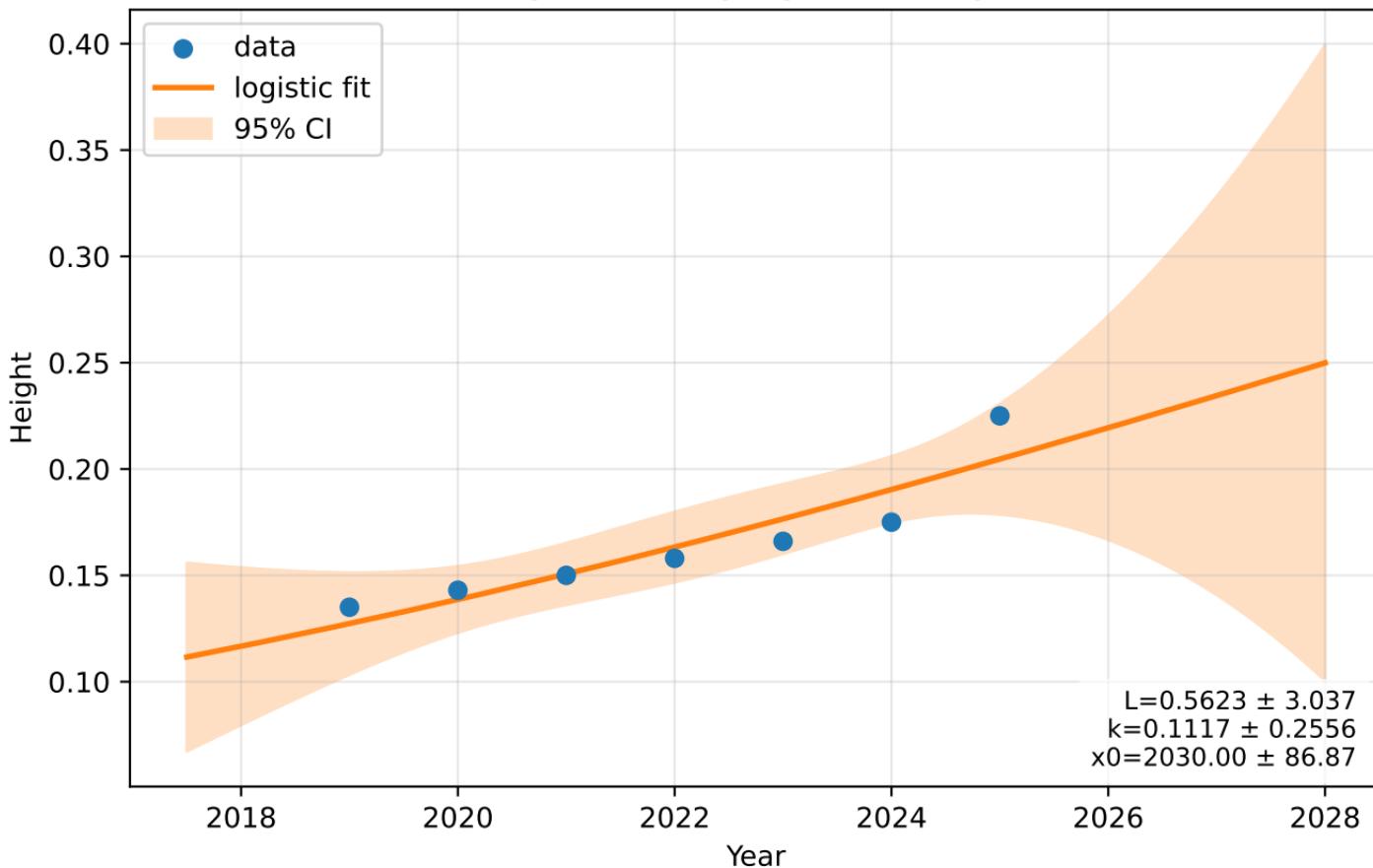
Complex Reasoning (complex_reasoning)



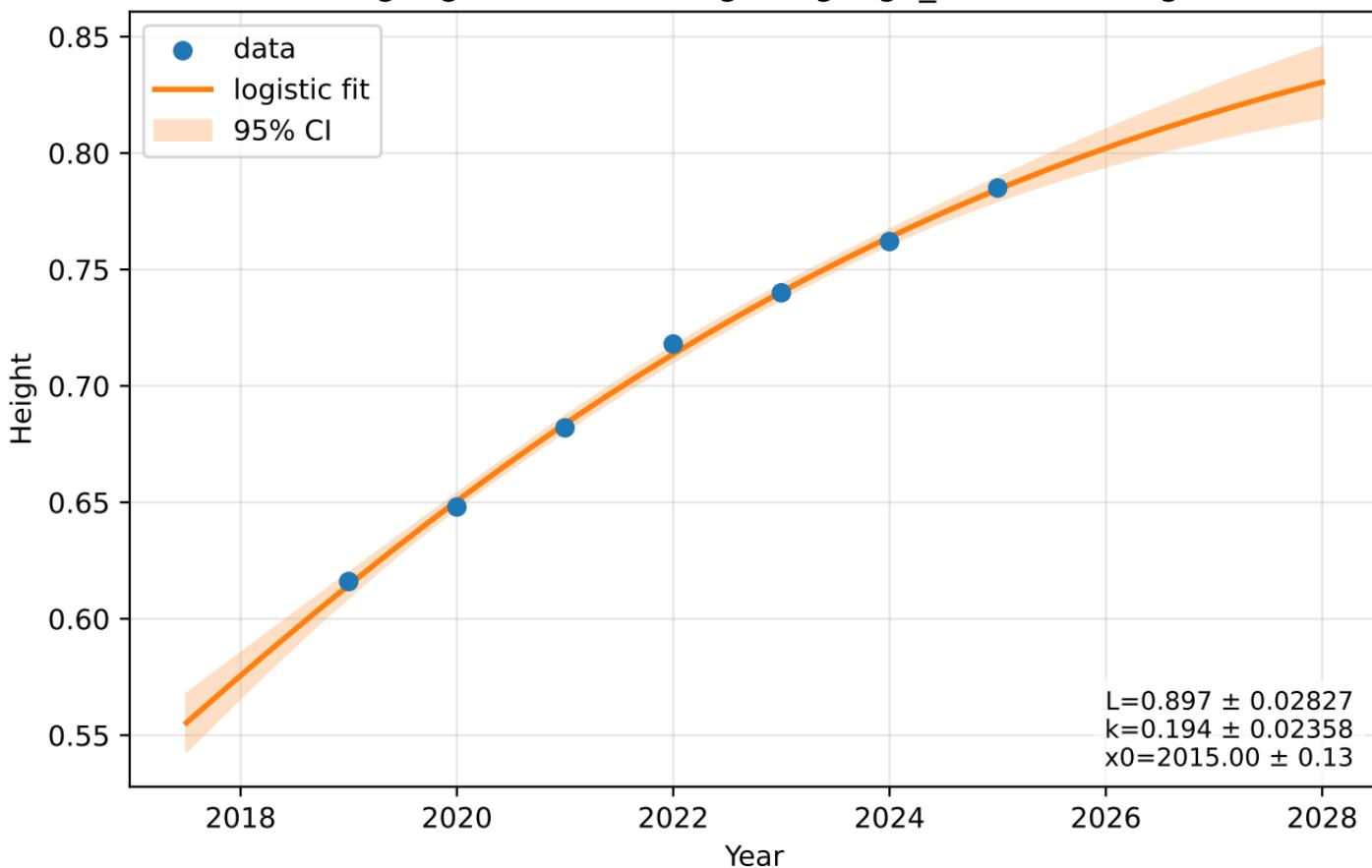
Spatial Reasoning (spatial_reasoning)



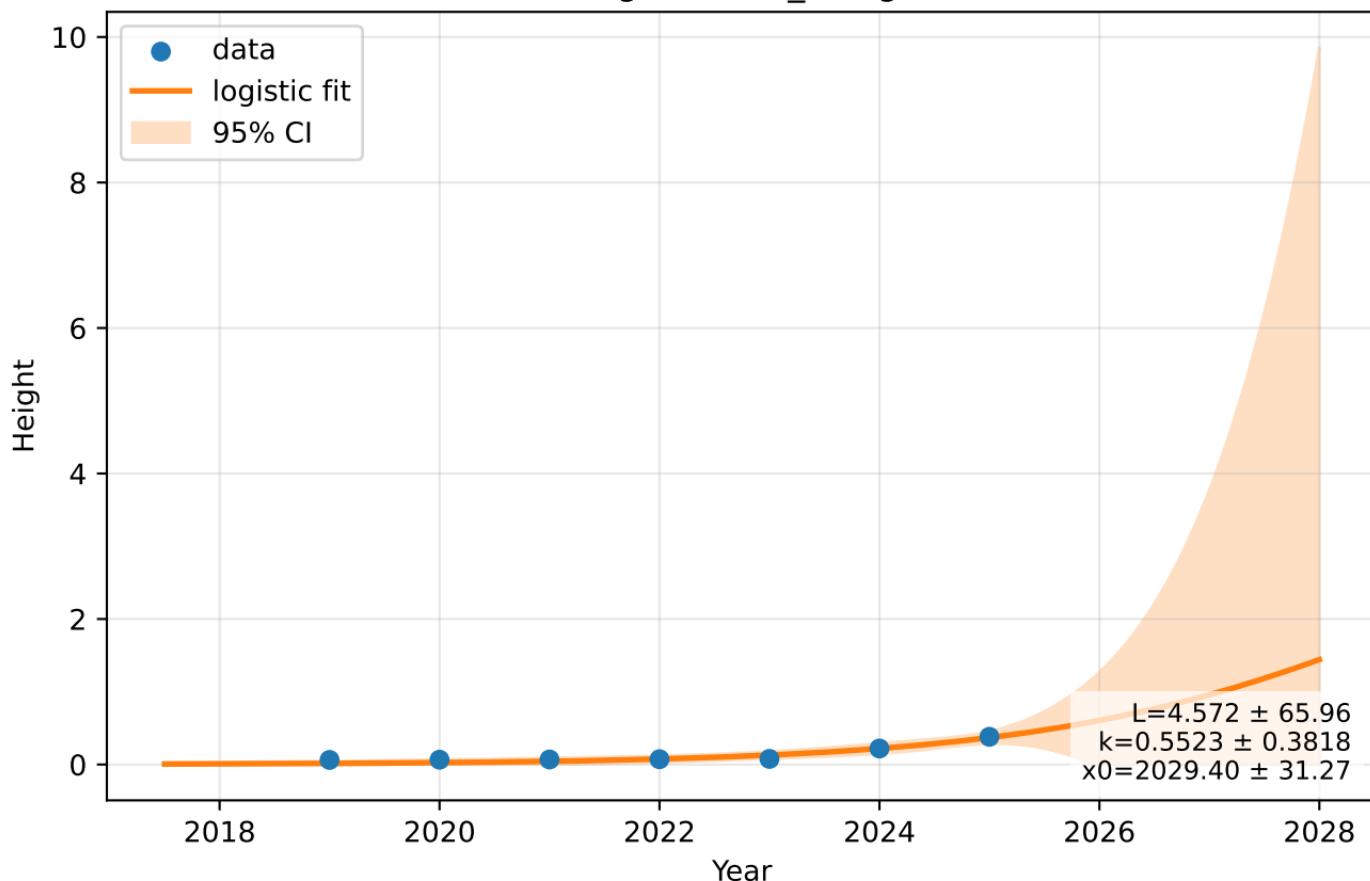
Cybersecurity (cybersecurity)



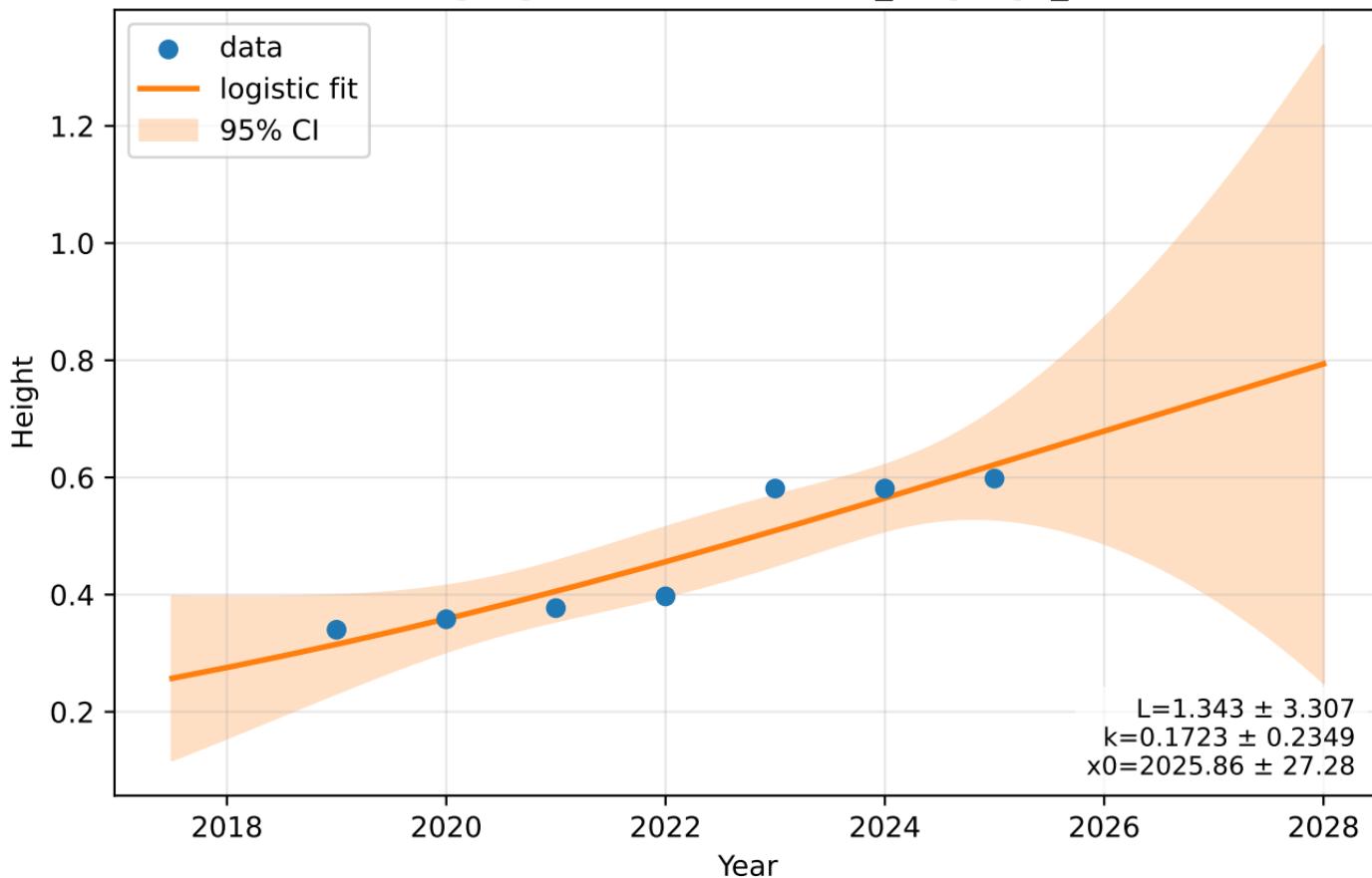
Language Understanding (language_understanding)



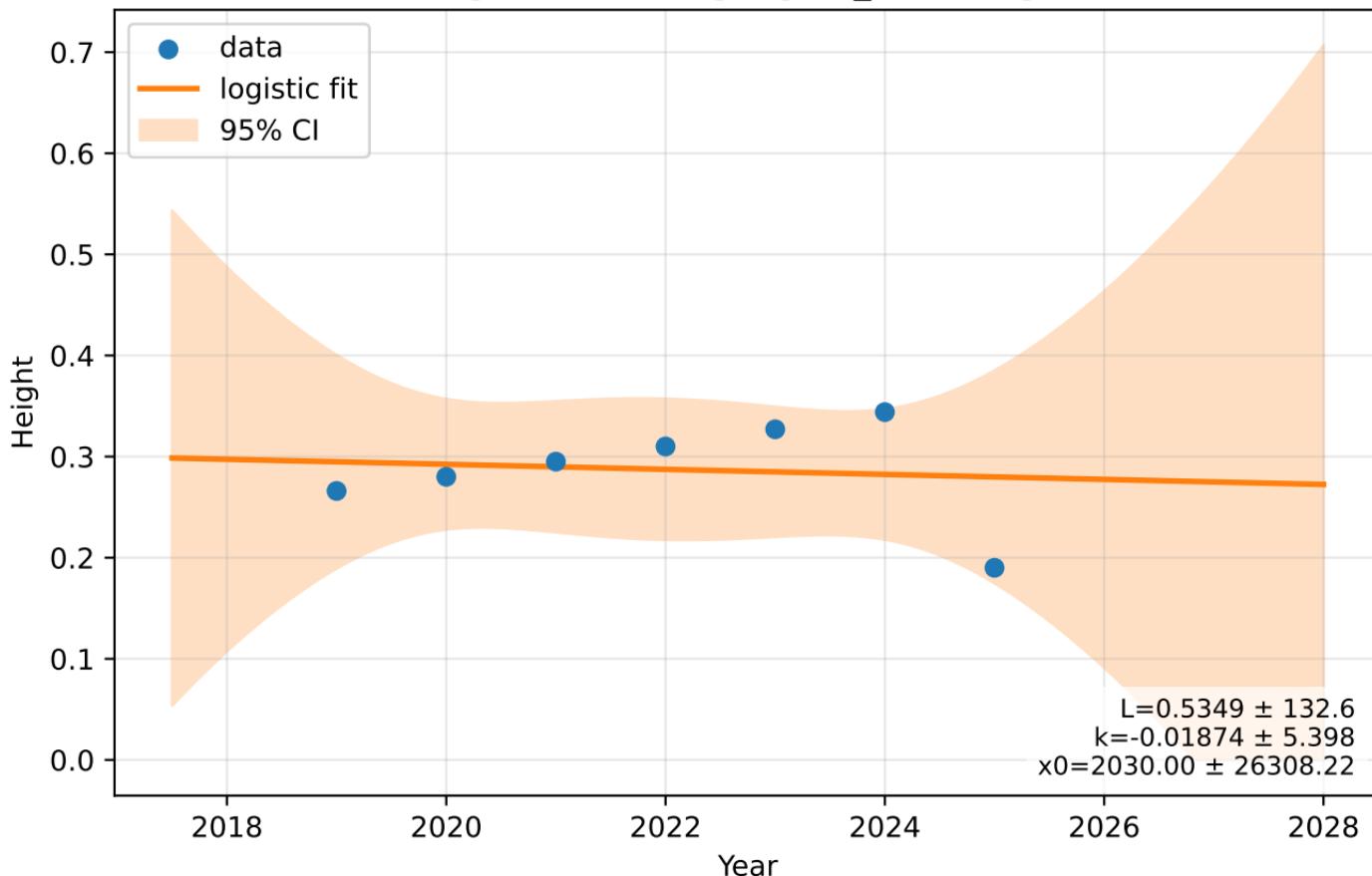
Os Navigation (os_navigation)



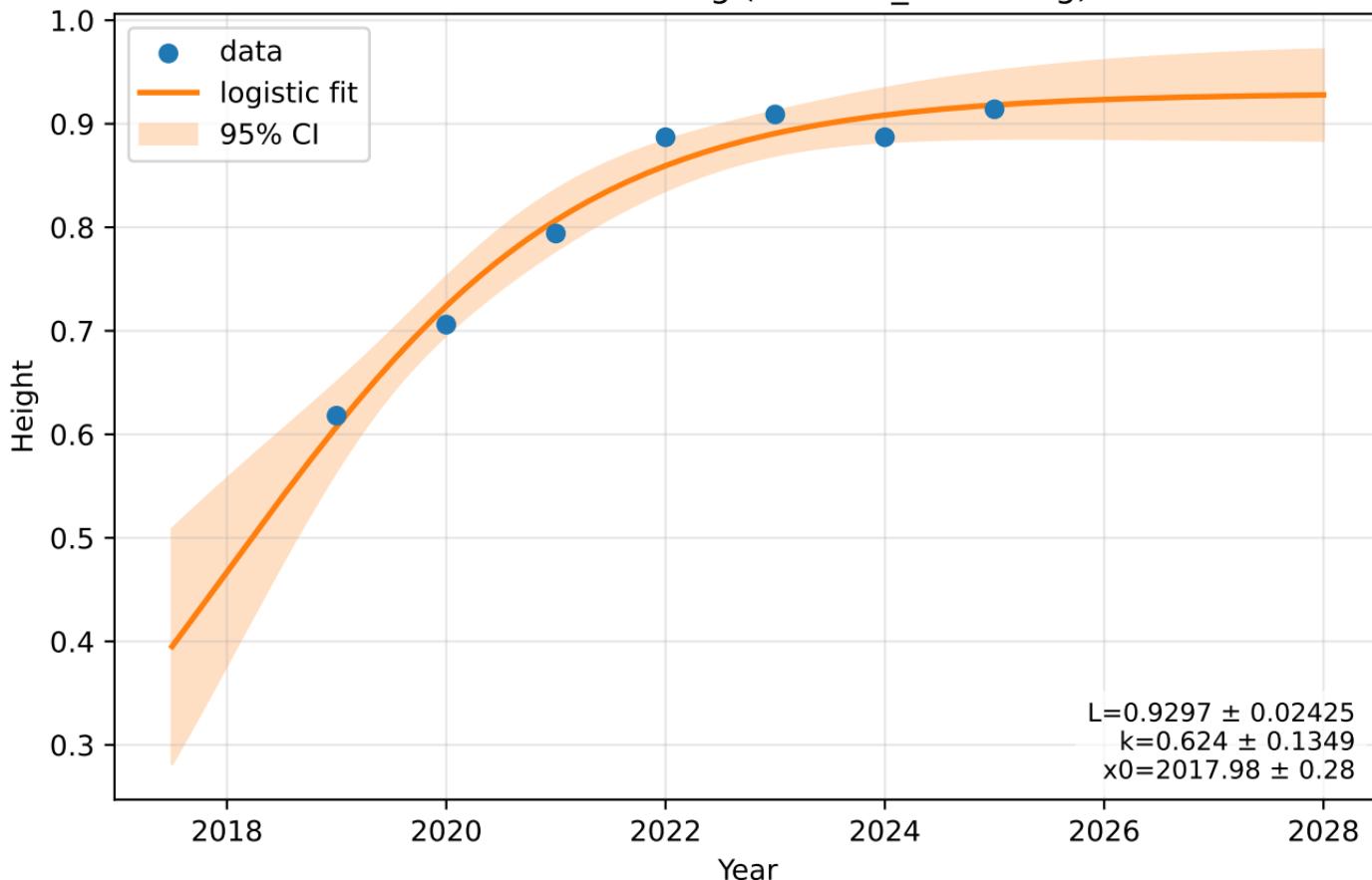
Natural Language Inference (natural_language_inference)



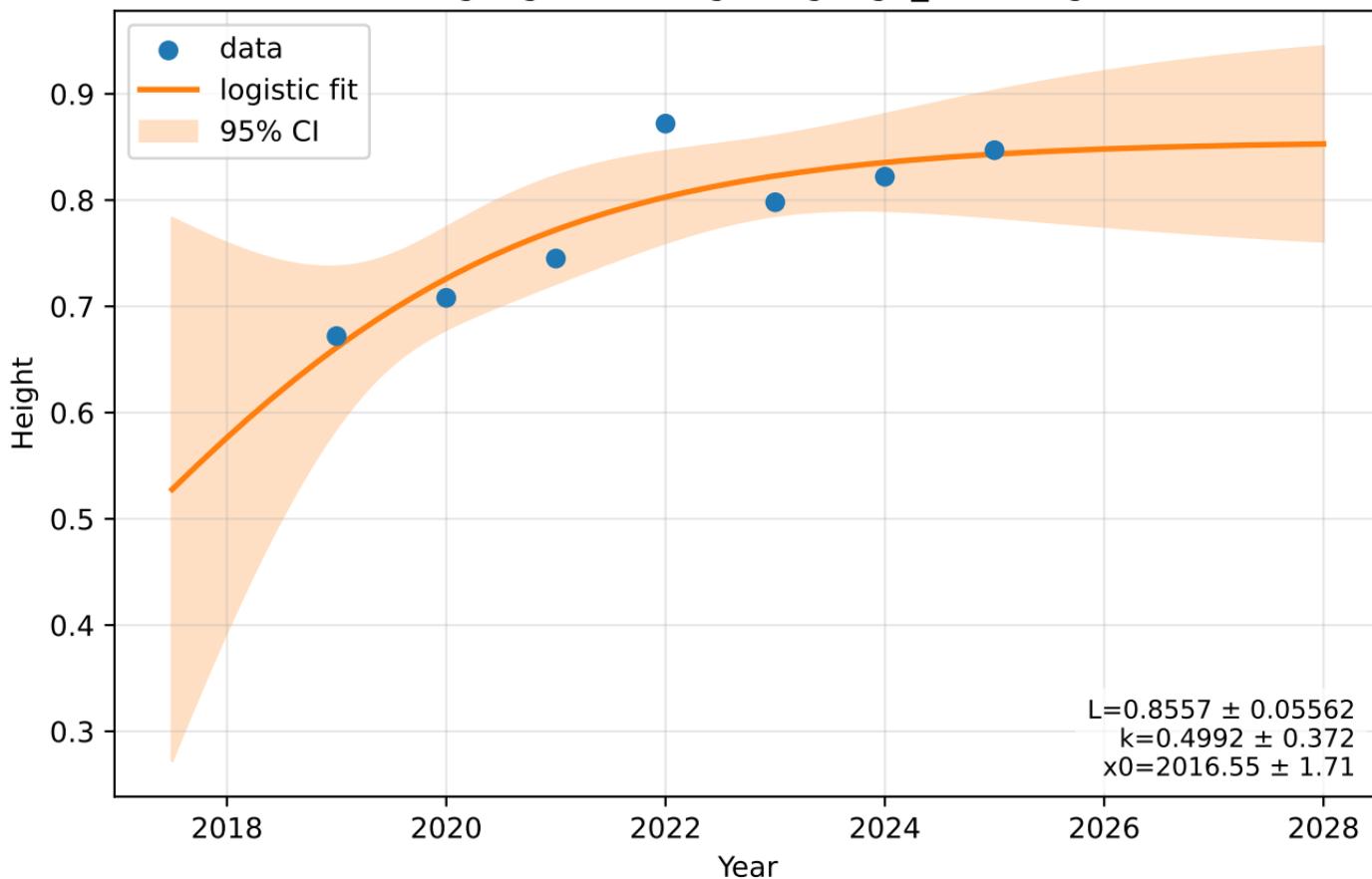
Agent Reasoning (agent_reasoning)



Boolean Reasoning (boolean_reasoning)



Language Modeling (language_modeling)



Terminal Usage (terminal_usage)

