

بسمه تعالی

موضوع پژوهش:

یادگیری تقویتی از بازخورد انسانی (RLHF)

بر اساس مقاله

"A Survey of Reinforcement Learning from Human Feedback" (Timo Kaufmann et al., 2024)

معرفی کلی پژوهش

این ریپوزیتوری، خلاصه‌ای ساختاریافته و تحلیل محتوای مقاله مرجع فوق است که حوزه "یادگیری تقویتی از بازخورد انسانی (RLHF)" را به طور جامع بررسی می‌کند. RLHF یک پارادایم قدرتمند در هوش مصنوعی است که در آن یک عامل، به جای تکیه بر یک تابع پاداش از پیش طراحی شده، هدف خود را مستقیماً از طریق تعامل با نظرات و ترجیحات انسان می‌آموزد. این روش نقشی کلیدی در همترازی (Alignment) سیستم‌های هوش مصنوعی با ارزش‌ها و اهداف بشری ایفا کرده و موفقیت چشمگیری در آموزش مدل‌های زبانی بزرگ مانند ChatGPT داشته است. در این پژوهش، با نگاهی فراتر از کاربردهای زبانی، به بررسی اصول پایه، معماری کلی، روش‌های جمع‌آوری داده، آموزش مدل پاداش و یادگیری سیاست در حوزه RLHF پرداخته شده است.

هدف و ارزش افزوده این کار

هدف این پروژه، ایجاد یک مرجع منسجم و به‌روز بر اساس مقاله مذکور برای پژوهشگران و توسعه‌دهندگانی است که قصد ورود به این حوزه پویا را دارند یا می‌خواهند دانش خود را در آن به روز نگه دارند. برخلاف بسیاری از مروارهای موجود که تنها بر روی یک جنبه (مانند تنظیم مدل‌های زبانی) متمرکز شده‌اند، این مقاله و در نتیجه این ریپوزیتوری با پوششی گسترده، تمامی مؤلفه‌های اصلی چارچوب RLHF از انواع مختلف بازخورد انسانی و روش‌های جمع‌آوری هوشمند آن گرفته تا مبانی نظری و معیارهای ارزیابی را در بر می‌گیرد.

این مجموعه نه تنها به عنوان یک نقشه راه برای درک فضای کلی پژوهش‌های RLHF عمل می‌کند، بلکه با دسته‌بندی سیستماتیک آخرین دستاوردها و اشاره به چالش‌های باز، بستری برای ایده‌پردازی و توسعه راه حل‌های نوآورانه در آینده فراهم می‌سازد.