제목: 영화 세대: 미디어 재단 모델 출연진
번역된 요약:
다음을 생성하는 기반 모델 출연진인 Movie Gen을 소개합니다 다양한 화면 비율과 동기화된 고품질 1080p HD 비디오 오디오. 또한
정확한 명령어 기반과 같은 추가 기능을 보여줍니다 사용자 이미지를 기반으로 개인화된 동영상을 편집하고 생성합니다. 우리 모델은 여러 작업에서
새로운 최첨단 기술을 설정합니다: 텍스트에서 비디오로 합성, 비디오 개인화, 비디오 편집, 비디오-투-audio 생성 및 텍스트에서 오디오로
생성됩니다. 가장 큰 비디오 생성 모델은 30B 매개 변수입니다 최대 컨텍스트 길이 73,000개의 비디오 토큰으로 학습된 트랜스포머, 초당
16프레임으로 16초의 생성된 비디오에 해당합니다. 우리는 아키텍처에 대한 여러 가지 기술 혁신과 단순화를 보여줍니다, 잠재 공간, 교육 목표
및 레시피, 데이터 큐레이션, 평가 다음을 가능하게 하는 프로토콜, 병렬화 기술 및 추론 최적화 사전 교육 데이터, 모델 크기 및 교육 확장의
이점을 누릴 수 있습니다 대규모 미디어 생성 모델을 훈련하기 위한 컴퓨팅. 본 논문이 연구 커뮤니티가 미디어의 발전과 혁신을 가속화할 수 있도록
지원합니다 세대 모델. 이 문서의 모든 동영상은 다음에서 시청할 수 있습니다
https://go.fb.me/MovieGenResearchVideos .
URL: https://huggingface.co/papers/2410.13720

제목: MixEval-X: 실제 데이터 혼합물의 모든 평가
번역된 요약:
다양한 양식을 인식하고 생성하는 것은 AI 모델이 다음을 수행하는 데 매우 중요합니다 실제 신호에서 효과적으로 학습하고 참여하려면 다음이 필요합니다 개발에 대한 신뢰할 수 있는 평가. 우리는 다음과 같은 두 가지 주요 문제를 식별합니다 현재 평가: (1) 일관성 없는 표준, 서로 다른 기준에 의해 형성됨 프로토콜과 성숙도 수준이 다양한 커뮤니티; 그리고 (2) 중요성 쿼리, 등급 및 일반화 편향. 이러한 문제를 해결하기 위해 다음을 소개합니다 최적화하도록 설계된 최초의 실제 벤치마크인 MixEval-X는 입력 및 출력 양식 전반에 걸친 평가를 표준화합니다.
우리는 제안합니다 다중 modal 벤치마크 혼합물 및 적응-수정 파이프라인은 다음과 같습니다 실제 작업 분포를 재구성하여 평가를 일반화합니다 실제 사용 사례에 효과적으로 적용됩니다. 광범위한 메타 평가를 통해 접근 방식은 벤치마크 샘플을 실제 작업과 효과적으로 정렬합니다 분포 및 모델 순위는 다음과 같은 분포와 밀접한 상관관계가 있습니다 크라우드소싱 실제 평가(최대 0.98). 포괄적인 서비스 제공 기존 모델과 조직의 순위를 재조정하고 인사이트를 제공하는 리더보드 다중 modal 평가에 대한 이해를 높이고 향후 연구에 대한 정보를 제공합니다.
URL: https://huggingface.co/papers/2410.13754

제목: MobA: 효율적인 모바일 작업 자동화를 위한 2단계 에이전트 시스템
번역된 요약:
현재 모바일 어시스턴트는 시스템 API에 대한 의존도가 제한되어 있습니다 복잡한 사용자 지침과 다양한 인터페이스로 인해 어려움을 겪고 있습니다 제한된 이해력과 의사 결정 능력. 이 문제를 해결하기 위해 멀티모달 기반의 새로운 휴대폰 에이전트인 MobA를 제안합니다 이해 및 계획 기능을 향상시키는 대형 언어 모델 정교한 2단계 에이전트 아키텍처를 통해. 고급 글로벌 에이전트(GA)는 사용자 명령을 이해하고 기록을 추적할 책임이 있습니다 기억 및 계획 작업. 하위 수준의 로컬 에이전트(LA)가 자세한 내용을 예측합니다 하위 tasks 및 메모리에 의해 안내되는 함수 호출 형태의 작업 GA. 반사 모듈을 통합하면 효율적인 작업을 완료할 수 있습니다 는 시스템이 이전에는 볼 수 없었던 복잡한 작업을 처리할 수 있도록 지원합니다. MobA는 다음을 증명합니다 작업 실행 효율성 및 완료율의 대폭적인 개선 MLLM 기반 모바일의 잠재력을 강조하는 실제 평가 어시스턴트.
URL: https://huggingface.co/papers/2410.13757

제목: 판사벤치: LLM 기반 판사 평가를 위한 벤치마크
번역된 요약:
LLM 기반 심사위원은 인간 평가의 확장 가능한 대안으로 부상했습니다 모델을 평가, 비교 및 개선하는 데 점점 더 많이 사용되고 있습니다.
하지만 LLM 기반 판사 자체의 신뢰성은 거의 조사되지 않습니다. LLM으로서 더욱 발전하고 대응이 더욱 정교해짐에 따라 다음과 같은 요구
사항이 발생합니다 더 강력한 심사위원이 평가합니다. 기존 벤치마크는 주로 판사가 인간의 선호도에 맞춰 조정하지만 종종 더 많은 것을 설명하지
못합니다 크라우드소싱된 인간 선호도가 다음과 같은 나쁜 지표인 어려운 작업 사실적이고 논리적인 정확성. 이 문제를 해결하기 위해 새로운 평가를
제안합니다 LLM 기반 심사위원을 객관적으로 평가하는 프레임워크입니다. 이 프레임워크를 기반으로 저희는 LLM 기반 심사위원의 도전적 평가를
위한 벤치마크인 JudgeBench 제안 지식, 추론, 수학, 코딩을 아우르는 응답 쌍. 저지벤치 새로운 파이프라인을 활용하여 기존의 어려운
데이터 세트를 다음과 같이 변환합니다 목표를 반영하는 선호도 레이블이 있는 어려운 응답 쌍 정확성. 신속한 심사위원 모음에 대한 종합적인 평가,
미세 조정된 심사위원, 다중 에이전트 심사위원 및 보상 모델에 따르면 JudgeBench는 다음과 같습니다 이전 벤치마크보다 훨씬 더 큰 도전
과제를 안고 있으며 무작위보다 약간 더 나은 성능을 발휘하는 강력한 모델(예: GPT-4o) 추측. 전반적으로 저지벤치는 신뢰할 수 있는 평가
플랫폼을 제공합니다 점점 더 발전하는 LLM 기반 심사위원. 데이터 및 코드는 다음에서 확인할 수 있습니다
https://github.com/ScalerLab/JudgeBench .
URL: https://huggingface.co/papers/2410.12784

제목: 대형 언어 모델을 사용한 초인적 음성 이해를 위한 로드맵
번역된 요약:
대형 언어 모델(LLM)의 성공으로 인해 통합하려는 노력이 활발해졌습니다 음성 및 오디오 데이터, 다음을 수행할 수 있는 일반적인 기초 모델 생성을 목표로 합니다 텍스트 입력과 비텍스트 입력을 모두 처리합니다 GPT-4o, 엔드투엔드 스피치 LLM의 잠재력을 강조하며 다음을 유지합니다 semantic가 아닌 정보와 세계 지식을 바탕으로 더 깊은 음성 이해를 할 수 있습니다. 음성 LLM의 개발을 안내하기 위해 5단계 로드맵을 제안합니다, 기본 자동 음성 인식(ASR)부터 고급 초인류까지 다양합니다 비 semantic 정보와 추상 음향을 통합할 수 있는 모델 복잡한 작업에 대한 지식. 또한 벤치마크인 SAGI Bechmark를 설계합니다, 이는 다음 5가지 수준의 다양한 작업에 걸쳐 중요한 측면을 표준화합니다, 추상적 음향 지식과 완전성을 사용하는 데 있어 어려움을 발견합니다 기능. 우리의 연구 결과는 병렬적 신호를 처리하는 데 있어 격차가 있음을 밝혀냈습니다 추상적인 음향 지식과 미래 방향을 제시합니다. 이 논문 음성 LLM을 발전시키기 위한 로드맵의 개요를 설명하고 다음을 위한 벤치마크를 소개합니다 평가하고 현재 한계에 대한 주요 인사이트를 제공합니다 잠재적인.
URL: https://huggingface.co/papers/2410.13268

제목: 야누스 통합 멀티모달 이해 및 생성을 위한 시각적 인코딩 분리하기
번역된 요약:
본 논문에서는 통합하는 자기 회귀 프레임워크인 야누스를 소개한다 다중 모드 이해 및 생성. 이전 연구에서는 카멜레온과 같은 두 작업 모두에 대한 단일 시각적 인코더. 그러나 멀티모달에 필요한 다양한 수준의 정보 세분화 이해와 생성, 이 접근 방식은 최적의 성능을 발휘하지 못할 수 있습니다, 특히 다중 모드 이해에서. 이 문제를 해결하기 위해 우리는 디커플링을 해제합니다 단일 경로를 활용하면서 별도의 경로로 시각적 인코딩, 처리를 위한 통합 트랜스포머 아키텍처. 디커플링뿐만 아니라 시각적 인코더의 이해 역할 간의 충돌을 완화합니다 생성뿐만 아니라 프레임워크의 유연성도 향상시킵니다. 예를 들어, 두 가지 모두 다중 모드 이해 및 생성 구성 요소는 독립적으로 선택할 수 있습니다 가장 적합한 인코딩 방법입니다. 실험에 따르면 야누스는 이전 통합 모델로 작업별 성능과 일치하거나 그 이상입니다 모델. 야누스의 단순성, 높은 유연성과 효과로 인해 차세대 통합 멀티모달 모델에 대한 강력한 후보입니다.
URL: https://huggingface.co/papers/2410.13848

제목: 텍스트가 풍부한 시각적 이해를 위한 웹페이지 UI 활용
번역된 요약:
텍스트가 풍부한 시각적 이해 - 다음과 같은 환경을 처리할 수 있는 기능 고밀도 텍스트 콘텐츠는 비주얼과 통합되어 멀티모달에 매우 중요합니다 구조화된 언어와 효과적으로 상호 작용하는 대형 언어 모델(MLLM) 환경. 이러한 기능을 강화하기 위해 일반적인 합성을 제안합니다 텍스트 기반 대형 언어 모델을 사용하는 웹페이지 UI의 멀티모달 명령어 (LLM). 직접적인 시각적 입력이 부족함에도 불구하고 텍스트 기반 LLM은 다음을 수행할 수 있습니다 웹페이지 접근성 트리에서 구조화된 텍스트 표현을 처리합니다 그런 다음 지침을 UI 스크린샷과 페어링하여 다중 모드 모델을 훈련합니다. 우리는 100만 개의 샘플 730만 개가 포함된 데이터 세트인 멀티UI를 소개합니다 다양한 멀티모달 작업과 UI 레이아웃을 다루는 웹사이트. 학습된 모델 멀티UI는 웹 UI 작업에서 탁월할 뿐만 아니라 최대 48₩% 향상된 성능을 발휘합니다 비주얼웹벤치 및 웹 에이전트 데이터 세트의 작업 정확도 19.1₩% 향상 마인드투웹은 웹이 아닌 UI 작업에도 놀라울 정도로 잘 일반화되어 있습니다 문서 이해, OCR, 차트 해석과 같은 비 UI 도메인. 이러한 결과는 발전을 위한 웹 UI 데이터의 광범위한 적용 가능성을 강조합니다 다양한 시나리오에서 텍스트가 풍부한 시각적 이해.
URL: https://huggingface.co/papers/2410.13824

제목: Fluid: 연속 토큰으로 자동 회귀 텍스트-이미지 생성 모델 확장
번역된 요약:
시력에서 자동 회귀 모델을 확장하는 것은 다음과 같이 유익한 것으로 입증되지 않았습니다 대형 언어 모델. 이 작업에서는 이 확장 문제를 조사합니다 텍스트-이미지 생성의 맥락, 두 가지 중요한 요소에 초점을 맞추고 있습니다 모델은 개별 토큰 또는 연속 토큰을 사용하며, 토큰의 생성 여부는 다음과 같습니다 BERT 또는 GPT와 유사한 트랜스포머 아키텍처를 사용하여 무작위 또는 고정 래스터 순서를 지정합니다. 경험적 결과에 따르면 모든 모델은 다음과 같은 측면에서 효과적으로 확장됩니다 검증 손실, 평가 성과 - FID, GenEval에서 측정합니다 점수 및 시각적 품질 - 다양한 트렌드를 따릅니다 연속 토큰은 다음을 사용하는 토큰보다 훨씬 더 나은 시각적 품질을 달성합니다 개별 토큰. 또한 생성 순서 및 주의 메커니즘 GenEval 점수에 상당한 영향을 미칩니다: 무작위 순서 모델은 눈에 띄게 달성합니다 래스터 주문 모델에 비해 GenEval 점수가 더 높습니다. 이에서 영감을 받았습니다 연구 결과, 연속적으로 무작위 순서 자기 회귀 모델인 Fluid를 훈련합니다 토큰. 플루이드 10.5B 모델, 6.16의 새로운 최첨단 제로 샷 FID 달성 MS-COCO 30K에서, GenEval 벤치마크에서 전체 점수 0.69점을 받았습니다. 저희가 연구 결과와 결과는 향후 더 많은 가교 역할을 하기 위한 노력을 장려할 것입니다 시각 및 언어 모델 간의 격차 확대.
URL: https://huggingface.co/papers/2410.13863

제목: 세계 요리: 글로벌 요리에 대한 다국어 및 다문화 시각적 질문 답변을 위한 대규모 벤치마크
번역된 요약:
비전 언어 모델(VLM)은 종종 문화별 지식으로 어려움을 겪습니다, 특히 영어 이외의 언어와 소외된 문화권에서 더욱 그러합니다 컨텍스트. 이러한 지식에 대한 이해를 평가하기 위해 다음을 소개합니다 다국어 및 다문화에 대한 대규모 벤치마크인 WorldCuisines, 시각적으로 근거한 언어 이해. 이 벤치마크에는 시각적 기능이 포함됩니다 30개 언어에 걸쳐 텍스트-이미지 쌍이 포함된 질문 응답(VQA) 데이터 세트와 9개 언어군에 걸쳐 있으며 100만 개 이상의 데이터가 포함된 방언 포인트, 역대 최대 규모의 다문화 VQA 벤치마크가 되었습니다. 여기에는 다음이 포함됩니다 요리 이름과 그 유래를 식별하기 위한 작업. 평가를 제공합니다 두 가지 크기(12,000개 및 60,000개 인스턴스)의 데이터 세트와 학습 데이터 세트(1개) 수백만 건). 연구 결과에 따르면 VLM은 다음과 같은 기능을 더 잘 수행하는 것으로 나타났습니다 올바른 위치 컨텍스트, 적대적 컨텍스트와 씨름하고 특정 지역 요리와 언어 예측. 미래를 지원하기 위해 연구 결과, 주석이 달린 식품 항목과 이미지가 포함된 지식 기반을 공개합니다 VQA 데이터와 함께.
URL: https://huggingface.co/papers/2410.12705

제목: 드림비디오-2: 정밀한 모션 제어를 통한 제로 샷 피사체 기반 비디오 커스터마이징
번역된 요약:
최근 맞춤형 비디오 생성의 발전으로 사용자는 다음과 같은 작업을 수행할 수 있게 되었습니다 특정 피사체와 움직임 궤적 모두에 맞춘 동영상.
하지만, 기존 방법은 종종 복잡한 테스트 시간 미세 조정이 필요하고 어려움을 겪습니다 피험자 학습과 동작 제어의 균형을 유지하여 실제 세계를
제한합니다 애플리케이션. 본 논문에서는 제로 샷 비디오인 드림비디오-2를 소개합니다 특정 주제로 동영상을 생성할 수 있는 사용자 지정 프레임워크
단일 이미지와 바운딩 박스 시퀀스에 의해 유도되는 움직임 궤적, 테스트 시간 미세 조정이 필요하지 않습니다. 특히 모델의 고유한 기능을 활용하는
참조 주의를 도입합니다 피험자 학습을 위한 기능과 다음을 위한 마스크 가이드 모션 모듈을 고안합니다 강력한 모션 신호를 최대한 활용하여 정확한
모션 제어를 달성합니다 바운딩 박스에서 파생된 박스 마스크. 이 두 구성 요소는 다음을 달성하는 동안 의도된 기능, 우리는 모션 제어가 다음과
같은 경향이 있음을 경험적으로 관찰합니다 주제 학습보다 우위를 점합니다. 이 문제를 해결하기 위해 두 가지 주요 설계를 제안합니다: 1) 혼합된
잠재 마스크 모델링을 통합하는 마스킹된 기준 주의력 참조 주의를 기울여서 피사체 표현을 향상시킵니다 원하는 위치 및 2) 가중치 확산 손실로
인해 경계 상자 내부와 외부 영역의 기여도를 보장합니다 피사체와 동작 제어 사이의 균형. 에 대한 광범위한 실험 결과 새로 선별된 데이터 세트는
드림비디오-2가 더 뛰어난 성능을 발휘한다는 것을 보여줍니다 피사체 커스터마이징과 모션 제어 모두에서 최첨단 방법. The 데이터 세트, 코드
및 모델은 공개될 예정입니다.
URL: https://huggingface.co/papers/2410.13830

제목: MMed-RAG: 의료 비전 언어 모델을 위한 다목적 멀티모달 RAG 시스템
번역된 요약:
인공 지능(AI)은 다음 분야에서 상당한 잠재력을 입증했습니다 특히 질병 진단 및 치료 계획에서 의료 서비스. 최근 의료용 대형 비전 언어 모델(Med-LVLM)의 발전이 새로운 지평을 열었습니다 대화형 진단 도구의 가능성. 그러나 이러한 모델은 종종 사실적 환각으로 인해 잘못된 진단으로 이어질 수 있습니다. 미세 조정 및 검색 증강 생성(RAG)이 다음을 수행하는 방법으로 부상했습니다 이러한 문제를 해결합니다. 그러나 고품질 데이터와 분포의 양 교육 데이터와 배포 데이터 간의 이동으로 인해 다음과 같은 적용이 제한됩니다 미세 조정 방법. RAG는 가볍고 효과적이지만, 기존 RAG 기반 접근 방식은 다양한 의료 영역에 충분히 일반적이지 않습니다 그리고 잠재적으로 양식 간의 정렬 오류 문제를 유발할 수 있습니다 모델과 실제 진실 사이에서. 본 논문에서는 다용도를 제안합니다 멀티모달 RAG 시스템, MMed-RAG는 다음과 같은 사실성을 향상시키도록 설계되었습니다 Med-LVLM. 우리의 접근 방식은 도메인 인식 검색 메커니즘인 적응형 검색 컨텍스트 선택 방법 및 입증 가능한 RAG 기반 선호도 미세 조정 전략. 이러한 혁신은 RAG 프로세스를 충분히 일반적이고 신뢰할 수 있으며, 다음과 같은 경우 정렬이 크게 개선됩니다 검색된 컨텍스트 소개. 5개 의료 분야에 걸친 실험 결과 의료용 VQA에 대한 데이터 세트(방사선학, 안과학, 병리학 포함) 및 보고서 생성을 통해 MMed-RAG가 평균 개선을 달성할 수 있음을 입증했습니다 Med-LVLM의 사실 정확도는 43.8%입니다. 데이터와 코드를 사용할 수 있습니다 https://github.com/richard-peng-xia/MMed-RAG 에서.
URL: https://huggingface.co/papers/2410.13085

제목: BenTo: 컨텍스트 내 전송 기능을 통한 작업 감소 벤치마크
번역된 요약:
대규모 언어 모델(LLM)을 평가하는 데는 비용이 많이 들기 때문에 세대가 필요합니다 그리고 다양한 작업의 대규모 벤치마크에서 LLM 출력을 검사합니다. 이 백서에서는 벤치마킹에 사용되는 작업을 효율적으로 줄이는 방법을 살펴봅니다 평가 품질에 영향을 미치지 않는 LLM. 우리의 연구에 따르면 다음과 같은 작업이 있습니다 이전 가능성 및 관련성은 가장 많은 것을 식별할 수 있는 중요한 정보를 제공합니다 시설 위치 함수 최적화를 통한 작업의 대표적인 하위 집합. 우리 전송 가능성을 추정하기 위한 실질적으로 효율적인 지표를 제안합니다 컨텍스트 내 학습(ICL)을 통해 두 작업 간. 쌍별 분석을 통해 전송 가능성, 최신 LLM 벤치마크(예: MMLU 또는 FLAN)에서 5%까지 평가에 4% 미만의 차이만 유도합니다 오리지널 벤치마크. 이전 작업과 비교했을 때, 우리의 방법은 훈련이 필요 없습니다, 기울기가 없고 ICL만 필요로 하는 고효율입니다.
URL: https://huggingface.co/papers/2410.13804

제목: MoH: 혼합 헤드 어텐션으로서의 멀티 헤드 어텐션
번역된 요약:
본 연구에서는 멀티헤드 어텐션 메커니즘의 핵심인 멀티헤드 어텐션 메커니즘을 업그레이드한다 트랜스포머 모델, 다음을 유지하거나 능가하면서 효율
개선합니다 이전 정확도 수준. 우리는 다중 헤드 주의력을 다음과 같이 표현할 수 있음을 보여줍니다 요약 양식. 모든 주의력을 집중시키는 것은
아니라는 통찰력을 바탕으로 동일한 의미에서, 우리는 새로운 혼합 머리 주의력(MoH)을 제안합니다 전문가 혼합물의 전문가로서 주의력을 집중시키는
아키텍처 (MoE) 메커니즘. MoH에는 두 가지 중요한 이점이 있습니다: 첫째, MoH는 다음을 모두 지원합니다 적절한 주의력 헤드를 선택하는
토큰으로 추론 효율성 향상 정확도를 떨어뜨리거나 매개변수 수를 늘리지 않고 말입니다. 둘째, MoH는 다중 헤드 어텐션에서 표준 합산을 가중치로
대체합니다 요약, 주의 메커니즘의 유연성 도입 및 잠금 해제 추가 성능 잠재력. ViT, DiT 및 LLM에 대한 광범위한 실험 MoH가
50~90%만 사용하여 멀티헤드 주의력을 능가한다는 것을 입증합니다 주의력 헤드. 또한 사전 훈련된 멀티헤드를 시연합니다 LLaMA3-8B와
같은 주의 모델은 MoH에 계속 조정할 수 있습니다 모델. 특히 MoH-LLaMA3-8B는 14개 항목에서 평균 64.0%의 정확도를
달성했습니다 벤치마크, 75%만 활용하여 LLaMA3-8B를 2.4% 초과 달성 주의하세요. 우리는 제안된 MoH가 다음과 같은 유망한
대안이라고 믿습니다 멀티헤드 주의를 기울이고 고급 개발을 위한 강력한 기반을 제공합니다 효율적인 주의력 기반 모델입니다.
URL: https://huggingface.co/papers/2410.11842

제목: PopAlign: 보다 포괄적인 정렬을 위한 대조 패턴 다양화
번역된 요약:
대형 언어 모델(LLM)의 정렬에는 다음에 대한 모델 교육이 포함됩니다 선호도-contrast 출력 쌍을 사용하여 다음에 따라 반응을 조정합니다
인간 선호도. 이러한 대조적인 쌍을 얻으려면 전통적인 방법은 다음과 같습니다 RLHF 및 RLAIF는 다양한 모델과 같은 제한된 대조 패턴에
의존합니다 변형 또는 디코딩 온도. 이 특이점은 두 가지 문제로 이어집니다: (1) 정렬이 포괄적이지 않으므로 (2) 모델은 다음에 취약합니다
탈옥 공격. 이러한 문제를 해결하기 위해 다음과 같은 방법을 조사합니다 선호도를 높이기 위한 보다 포괄적이고 다각화된 대조 패턴 데이터(RQ1)
및 대조 패턴의 다양화가 미치는 영향 확인 모델 정렬(RQ2)에서. RQ1의 경우, 다음과 같은 프레임워크인 PopAlign을 제안합니다
프롬프트, 모델, 그리고 다양한 대조 패턴을 통합합니다 파이프라인 레벨, 필요 없는 6가지 대조 전략 도입 추가 피드백 라벨링 절차. RQ2와
관련하여 철저한 조사를 실시합니다 팝얼라인이 기존보다 훨씬 뛰어난 성능을 발휘한다는 실험 결과가 나왔습니다 보다 포괄적인 정렬로 이어지는 방법
URL: https://huggingface.co/papers/2410.13785

제목: 훈련 후 대규모 모델에서 델타 매개변수 편집에 대한 통합된 관점
번역된 요약:
사후 교육은 대규모 적응을 위한 중요한 패러다임으로 부상했습니다 델타에 의해 효과가 완전히 반영되는 다양한 작업에 대한 사전 학습된 모델 매개변수(즉, 사전 학습된 매개 변수와 사후 학습된 매개 변수 간의 차이) 매개변수). 수많은 연구에서 델타 매개변수 속성을 탐구했지만 가지치기, 양자화, 낮은 순위 근사화 등의 연산을 통해 외삽, 이를 체계적으로 검토하기 위한 통합 프레임워크 특성이 부족했습니다. 본 논문에서는 새로운 관점을 제안합니다 델타를 설명하기 위한 손실 함수의 리만 합 근사치를 기반으로 합니다 매개 변수 편집 작업. 저희 분석은 기존 방법을 다음과 같이 분류합니다 editing 후 성과에 따른 세 가지 수업: 경쟁, 감소, 리만 합으로 표현되는 방식을 설명하며 개선되었습니다 근사 항과 모델 성능을 변경하는 방법. 광범위 ViT, LLaMA 3, Qwen 2를 포함한 시각 및 언어 모델에 대한 실험, 그리고 미스트랄은 우리의 이론적 발견을 확증합니다. 또한 다음과 같이 소개합니다 DARE 및 비트델타와 같은 기존 기술을 확장하여 강조합니다 델타 매개변수의 속성을 활용하고 재구성하는 데 있어 제한 사항 적용 가능성과 효과를 높이기 위해 일반적인 표현으로 사용합니다 학습 후 모델에서 델타 매개 변수 편집.
URL: https://huggingface.co/papers/2410.13841

제목: OpenAI의 o1 모델 추론 패턴 비교 연구
번역된 요약:
대형 언어 모델(LLM)이 더 광범위한 복잡성을 처리할 수 있도록 지원 과제(예: 코딩, 수학)는 많은 연구자들로부터 큰 관심을 받고 있습니다 LLM은 모델 매개변수의 수를 늘리는 데 그치지 않고 계속 진화하고 있습니다 는 성능 향상과 막대한 계산 비용을 감소시킵니다. 최근 OpenAI의 o1 모델은 추론 전략(즉, 테스트 시간 계산 방법)도 추론을 크게 향상시킬 수 있습니다 LLM의 기능. 그러나 이러한 방법의 메커니즘은 여전히 미개척. 우리 작업에서 O1의 추론 패턴을 조사하기 위해 우리는 o1을 기존 테스트 시간 계산 방법과 비교합니다(BoN, 단계별 BoN, 에이전트) 워크플로 및 셀프 리파인)을 통해 일반적으로 OpenAI의 GPT-4o를 백본으로 사용합니다 세 가지 영역(즉, 수학, 코딩, 상식)의 추론 벤치마크 추론). 구체적으로, 첫째, 우리의 실험에 따르면 O1 모델은 다음과 같은 기능을 가지고 있습니다 대부분의 데이터 세트에서 최고의 성능을 달성했습니다. 둘째 다양한 응답(예: BoN)을 검색하면 보상 모델의 기능을 찾을 수 있습니다 검색 공간은 모두 이러한 방법의 상한을 제한합니다. 셋째 문제를 여러 하위 문제로 나누는 방법, 에이전트 워크플로는 다음과 같은 이유로 단계별 BoN보다 더 나은 성능을 달성했습니다 더 나은 추론 프로세스를 계획하기 위한 도메인별 시스템 프롬프트. 넷째, 우리는 o1의 6가지 추론 패턴을 요약했다는 점을 언급할 가치가 있습니다, 는 여러 추론 벤치마크에 대한 자세한 분석을 제공했습니다.
URL: https://huggingface.co/papers/2410.13639

제목: VidPanos: 캐주얼 패닝 동영상의 제너레이티브 파노라마 동영상
번역된 요약:
파노라마 이미지 스티칭은 다음과 같은 장면을 통합된 광각으로 볼 수 있습니다 카메라의 시야 밖으로 확장됩니다. 패닝 비디오의 프레임 스티칭 파노라마 사진으로의 전환은 정지된 장면에서 잘 알려진 문제입니다, 하지만 물체가 움직일 때는 정지된 파노라마로는 장면을 포착할 수 없습니다. 우리는 캐주얼 captured에서 파노라마 동영상을 합성하는 방법을 제시합니다 마치 광각 카메라로 원본 비디오를 캡처한 것처럼 패닝하는 동영상입니다. 우리는 파노라마 합성을 시공간 아웃페인팅 문제로 상정하고, 여기서 목표로 합니다 입력 동영상과 동일한 길이의 전체 파노라마 동영상을 만듭니다. 일관성 있음 시공간 볼륨을 완성하려면 다음과 같은 강력하고 현실적인 사전 작업이 필요합니다 비디오 콘텐츠와 모션, 여기에 제너레이티브 비디오 모델을 적용합니다. 기존 그러나 생성 모델이 파노라마 완성으로 즉시 확장되는 것은 아닙니다, 우리가 보여주는 것처럼. 대신 비디오 생성을 파노라마의 구성 요소로 적용합니다 합성 시스템, 모델의 강점을 활용하는 방법 시연 한계를 최소화하면서. 저희 시스템은 사람, 차량, 흐르는 물을 포함한 다양한 야생 장면 고정된 배경 기능도 있습니다.
URL: https://huggingface.co/papers/2410.13832

제목: 플랫퀀트: LLM 양자화에 중요한 플랫퀀트
번역된 요약:
최근 양자화는 압축에 널리 사용되고 있습니다 대형 언어 모델의 가속화~(LLM). LLM의 이상값으로 인해 양자화 오류를 최소화하기 위해 가중치와 활성화를 평탄화하는 데 매우 중요합니다 동일한 간격의 양자화 포인트를 사용합니다. 이전 연구에서는 다양한 방법을 탐구합니다 채널별 등 이상치를 억제하기 위한 사전 양자화 변환 스케일링 및 하다마드 변환. 그러나 우리는 이것이 변환되었음을 관찰합니다 가중치와 활성화는 여전히 가파르고 넓게 유지될 수 있습니다. 이 논문에서 우리는 플랫퀀트(빠르고 학습 가능한 아핀 변환)를 새롭게 제안합니다 웨이트의 평탄성을 향상시키는 훈련 후 양자화 접근 방식 및 활성화. 우리의 접근 방식은 다음에 맞는 최적의 아핀 변환을 식별합니다 각 선형 레이어는 경량 목표를 통해 몇 시간 단위로 보정됩니다 런타임 오버헤드, 변환에 크로네커 분해를 적용합니다 행렬 및 플랫퀀트의 모든 연산을 단일 커널에 융합합니다. 광범위 실험에 따르면 플랫퀀트는 새로운 최첨단 양자화를 설정합니다 벤치마크. 예를 들어, 다음과 같은 경우 1% 미만의 정확도 저하를 달성합니다 LLaMA-3-70B 모델에서 스핀퀀트를 다음과 같이 능가하는 W4A4 양자화 7.5%. 추론 지연 시간의 경우 FlatQuant은 유도된 속도 저하를 줄입니다 0.26배의 QuaRot에서 단순히 양자화 전 변환을 통해 0.07배, 프리필 속도 최대 2.3배 향상 및 디코딩 속도가 각각 1.7배 빨라집니다. 코드는 다음에서 사용할 수 있습니다: https://github.com/ruikangliu/FlatQuant .
URL: https://huggingface.co/papers/2410.09426

제목: MedMobile: 전문가 수준의 임상 역량을 갖춘 모바일 크기의 언어 모델
번역된 요약:
언어 모델(LM)은 전문가 수준의 추론과 회상을 입증했습니다 의학 분야의 능력. 하지만 계산 비용과 개인정보 보호 문제는 다음과 같습니다 광범위한 구현에 대한 장벽이 높아지고 있습니다. 간결한 방법을 소개합니다 Phi-3-mini, MedMobile의 적응, 38억 개의 매개변수 LM, 다음을 수행할 수 있습니다 의료 애플리케이션을 위해 모바일 장치에서 실행됩니다 MedMobile은 MedQA(USMLE)에서 75.7%의 점유율을 기록하며 합격점을 넘어섰습니다 의사(~60%), 크기의 100배에 달하는 모델 점수에 근접했습니다. 우리는 그런 다음 신중한 절제 세트를 수행하고 다음과 같은 사슬을 시연합니다 사고, 조립, 미세 조정은 가장 큰 성능 향상으로 이어집니다, 예기치 않게 검색한 증강 생성은 다음을 증명하지 못합니다 상당한 개선 사항
URL: https://huggingface.co/papers/2410.09019

제목: 상호 작용을 통한 회고적 학습
번역된 요약:
대형 언어 모델(LLM)과 사용자 간의 멀티 턴 상호 작용 암시적 피드백 신호가 자연스럽게 포함됩니다. LLM이 다음과 같은 방식으로 응답하는 경우 예기치 않은 명령어의 방법, 사용자는 다시 표현하여 신호를 보낼 가능성이 높습니다 요청, 불만 표출 또는 대체 작업으로 전환하기. 그러한 신호는 작업에 독립적이며 상대적으로 제한된 하위 공간을 차지합니다 언어, LLM이 실제에서 실패하더라도 식별할 수 있도록 허용 작업. 이를 통해 다음과 같은 요소 없이 상호 작용을 통해 지속적으로 학습할 수 있는 방법이 만들어집니다 추가 주석. 이를 통해 학습할 수 있는 방법인 ReSpect를 소개합니다 회고를 통해 과거 상호작용의 신호를 보냅니다. 우리는 새로운 방식으로 리스펙트를 배포합니다 인간이 LLM에게 다음을 해결하도록 지시하는 다중 모드 상호 작용 시나리오 조합 솔루션 공간이 있는 추상적 추론 작업. 수천 개의 작업을 통해 인간과의 상호 작용에 대해 리스펙트가 점진적으로 작업을 개선하는 방법을 보여줍니다 31%에서 82%의 완료율을 기록했으며, 모두 외부 주석 없이 완료되었습니다.
URL: https://huggingface.co/papers/2410.13852

제목: 실패한 미래: 합성 데이터 및 검색 증강을 통해 ASR의 생성 오류 수정 개선
번역된 요약:
생성 오류 수정(GEC)은 강력한 사후 처리로 부상했습니다 자동 음성 인식(ASR)의 성능을 향상시키는 방법 시스템. 그러나 GEC 모델이 다음을 넘어 일반화하는 데 어려움을 겪고 있음을 보여줍니다 교육 중에 발생하는 특정 유형의 오류로 인해 다음과 같은 기능이 제한됩니다 테스트 시점, 특히 도메인 밖(OOD)에서 보이지 않는 새로운 오류 수정 시나리오. 이 현상은 명명된 개체(NE)로 증폭되며, 여기서 는 다음과 같습니다 NE에 대한 컨텍스트 정보나 지식이 충분하지 않을 뿐만 아니라, 새로운 NE가 계속 등장하고 있습니다. 이러한 문제를 해결하기 위해 DARAG(데이터 및 검색-증강 생성 오류 수정)을 위해 설계된 새로운 접근 방식 도메인 내(ID) 및 OOD 시나리오에서 ASR에 대한 GEC를 개선합니다. GEC를 보강합니다 LLM을 프롬프트하여 생성된 합성 데이터로 데이터 세트를 교육합니다 텍스트에서 speech로 모델을 변환하여 다음과 같은 추가 오류를 시뮬레이션합니다 모델은 학습할 수 있습니다. OOD 시나리오의 경우, 새로운 테스트 시간 오류를 시뮬레이션합니다 유사하게 비지도 방식으로 도메인을 관리합니다. 또한 더 나은 방법으로 명명된 개체를 처리하고 다음과 같은 방법으로 검색-augmented 수정을 도입합니다 데이터베이스에서 검색된 개체로 입력을 보강합니다. 우리의 접근 방식은 다음과 같습니다 간단하고 확장 가능하며 도메인과 언어에 구애받지 않습니다. 우리는 여러 데이터 세트와 설정, DARAG가 모든 데이터 세트를 능가한다는 것을 보여줍니다 기준선, ID에서 8₩% -- 30₩%의 상대적 WER 개선 및 10₩% 달성 -- 33₩%의 OOD 설정 개선.
URL: https://huggingface.co/papers/2410.13198

제목: 기억하고, 검색하고, 생성하기: 개인화된 어시스턴트로서 무한 시각적 개념 이해하기
번역된 요약:
대형 언어 모델(LLM)의 개발은 크게 향상되었습니다 멀티모달 LLM(MLLM)의 일반 어시스턴트로서의 기능. 하지만, 사용자별 지식 부족으로 인해 여전히 인간의 삶에 적용하는 데 한계가 있습니다 일상 생활. 본 논문에서는 검색 증강 개인화를 소개합니다 (RAP) MLLM의 개인화 프레임워크. 일반 MLLM부터 시작하여 세 단계로 개인화된 어시스턴트로 전환합니다. (a) 기억하세요: 저희는 사용자 이름, 아바타 등 사용자 관련 정보를 저장하는 키 값 데이터베이스 기타 속성. (b) 검색: 사용자가 대화를 시작하면 RAP 는 멀티모달을 사용하여 데이터베이스에서 관련 정보를 검색합니다 리트리버. (c) 생성: 입력 쿼리 및 검색된 개념의 정보 는 MLLM에 입력되어 개인화된 지식 증강 응답을 생성합니다. 이전 방법과 달리 RAP는 업데이트를 통해 실시간 개념 편집을 가능하게 합니다 외부 데이터베이스. 생성 품질 및 정렬을 더욱 개선하려면 사용자별 정보, 데이터 수집을 위한 파이프라인을 설계하고 다음을 생성합니다 MLLM의 개인화된 교육을 위한 특수 데이터 세트입니다. 데이터 세트를 기반으로, 우리는 일련의 MLLM을 개인화된 멀티모달 어시스턴트로 훈련시킵니다. By 대규모 데이터 세트에 대한 사전 학습, RAP-MLLM은 무한 시각으로 일반화할 수 있습니다 추가적인 미세 조정이 없는 개념. 우리 모델은 뛰어난 성능을 입증합니다 다음과 같은 다양한 작업에 걸친 유연성 및 생성 품질 개인화된 이미지 캡션, 질문 답변 및 시각 인식. 더 코드, 데이터 및 모델은
https://github.com/Hoar012/RAP-MLLM 에서 확인할 수 있습니다.
URL: https://huggingface.co/papers/2410.13360

제목: γ-$\MoD$: 멀티모달 대형 언어 모델을 위한 혼합 심층 적응 살펴보기
번역된 요약:
멀티모달 대형 언어 모델(MLLM)의 상당한 발전에도 불구하고, 높은 계산 비용은 여전히 실제 배포의 장벽으로 남아 있습니다. 자연어 처리의 깊이(MoD) 혼합에서 영감을 받아 목표를 달성했습니다 "활성화된 토큰"의 관점에서 이러한 한계를 해결합니다. 우리의 핵심 인사이트는 대부분의 토큰이 레이어 계산을 위해 중복되는 경우, 그런 다음 MoD 레이어를 통해 직접 건너뛸 수 있습니다. 그러나 직접 변환 MLLM에서 MoD 레이어로 구성된 조밀한 레이어는 상당한 성능을 제공합니다 저하. 이 문제를 해결하기 위해 혁신적인 MoD 적응을 제안합니다 감마-MoD라고 하는 기존 MLLM에 대한 전략. 감마-MoD에서, 소설 MLLM에서 MoD의 배포, 즉 다음과 같은 순위를 안내하기 위해 메트릭이 제안됩니다 주의 지도(ARANK). ARANK를 통해 어떤 레이어를 효과적으로 식별할 수 있습니다 중복되므로 MoD 계층으로 대체해야 합니다. ARank를 기반으로 우리는 계산 희소성을 극대화하기 위해 두 가지 새로운 설계를 추가로 제안합니다 MLLM은 성능을 유지하면서 비전 언어 공유기를 공유합니다 그리고 마스킹 라우팅 학습. 이러한 설계를 통해 90% 이상의 밀도가 높은 레이어를 MLLM은 MoD로 효과적으로 변환할 수 있습니다. 우리의 방법을 검증하기 위해, 우리는 이를 세 가지 인기 있는 MLLM에 적용하고 9에 대해 광범위한 실험을 수행합니다 벤치마크 데이터 세트. 실험 결과는 중요성을 검증할 뿐만 아니라 감마-MoD가 기존 MLLM에 미치는 효율성 이점뿐만 아니라 그 효과도 확인했습니다 다양한 MLLM의 일반화 기능. 예를 들어, 약간의 성능으로 drop, 즉 -1.5% 감마-MoD는 다음과 같은 훈련 및 추론 시간을 줄일 수 있습니다 LlaVA-HR은 각각 31.0%와 53.2% 감소했습니다.
URL: https://huggingface.co/papers/2410.13859

제목: MLLM은 중국 이미지의 깊은 영향을 이해할 수 있습니까?
번역된 요약:
멀티모달 대형 언어 모델(MLLM)의 기능이 계속 증가함에 따라 개선, MLLM의 고차 능력 평가의 필요성은 다음과 같습니다 증가하고 있습니다. 그러나 고차에 대한 MLLM을 평가하는 작업이 부족합니다 중국 시각 콘텐츠에 대한 인식과 이해. 그 공백을 메우기 위해 우리는 **C**hinees **I**이미지 소개 **복사 이해하기 **Bench**mark, **CII-Bench**, 고차원적 인식 평가를 목표로 합니다 중국 이미지에 대한 MLLM의 기능 이해하기. CII-Bench 스탠드 기존 벤치마크와 비교하여 여러 가지 방식으로 출시됩니다. 첫째 중국어 문맥의 진위 여부, CII-Bench의 이미지는 중국 인터넷 및 수동 검토, 해당 답변도 제공 수작업으로 제작되었습니다. 또한 CII-Bench는 다음과 같은 이미지를 통합합니다 유명한 중국 전통 그림과 같은 중국 전통 문화, 중국 전통에 대한 모델의 이해를 깊이 반영할 수 있습니다 문화. 여러 MLLM에 걸쳐 CII-Bench에 대한 광범위한 실험을 통해 는 중요한 발견을 했습니다. 처음에는 상당한 격차가 관찰됩니다 CII-Bench에서 MLLM과 인간의 성능 사이. 가장 높은 정확도 MLLM의 경우 64.4%에 도달하는 반면, 사람의 정확도는 평균 78.2%로 다음과 같은 수준에서 정점을 찍습니다 인상적인 81.0%. 그 후 MLLM은 중국 전통 제품보다 성능이 더 떨어집니다 문화 이미지, 이해 능력의 한계를 시사합니다 중국 전통에 대한 깊은 지식 기반이 부족하고 높은 수준의 의미론 문화. 마지막으로, 대부분의 모델은 향상된 정확도를 보이는 것으로 관찰됩니다 이미지 감정 힌트가 프롬프트에 통합될 때. 저희는 다음과 같이 믿습니다 CII-Bench를 통해 MLLM은 중국어 의미를 더 잘 이해할 수 있습니다 전문가 인공지능을 향한 여정을 발전시키는 중국 전용 이미지 및 중국 전용 이미지 일반 인텔리전스(AGI). 우리 프로젝트는 다음에서 공개적으로 이용할 수 있습니다 https://cii-bench.github.io/ .
URL: https://huggingface.co/papers/2410.13854

제목: Long-LRM: 넓은 범위의 가우스 스플랫을 위한 긴 시퀀스 대형 재구성 모델
번역된 요약:
우리는 다음과 같은 일반화 가능한 3D 가우시안 재구성 모델인 Long-LRM을 제안한다 긴 입력 이미지 시퀀스에서 큰 장면을 재구성할 수 있습니다. 특히, 저희 모델은 960x540 해상도로 32개의 소스 이미지를 처리할 수 있습니다 단일 A100 80G GPU에서 단 1.3초 이내에. 저희 아키텍처는 최근 맘바2 블록과 클래식 변압기 블록의 혼합물은 다음과 같습니다 이전 작업보다 더 많은 토큰을 처리할 수 있었으며, 효율성이 향상되었습니다 토큰 병합 및 가우시안 가지치기 단계는 품질과 가우시안 가지치기 사이의 균형을 유지합니다 효율성. 처리에 국한된 이전 피드 포워드 모델과 달리 1~4개의 입력 이미지로 큰 장면의 일부만 재구성할 수 있습니다, Long-LRM은 단일 피드 포워드 단계에서 전체 장면을 재구성합니다. On DL3DV-140 및 탱크 앤 템플과 같은 대규모 장면 데이터 세트, 우리의 방법 최적화 기반 접근 방식과 동등한 성능을 달성하는 동시에 두 가지 정도 더 효율적입니다. 프로젝트 페이지:
https://arthurhero.github.io/projects/llrm
URL: https://huggingface.co/papers/2410.12781

제목: 오픈 머티리얼즈 2024(OMAT24) 무기 재료 데이터 세트 및 모델
번역된 요약:
바람직한 특성을 가진 신소재를 발견하는 능력은 매우 중요합니다 기후 변화 완화에 도움이 되는 것부터 다음과 같은 발전에 이르기까지 다양한 응용 분야에 적용됩니다 차세대 컴퓨팅 하드웨어. AI는 가속화할 수 있는 잠재력을 가지고 있습니다 화학 공간을 보다 효과적으로 탐색하여 재료를 발견하고 설계합니다 다른 계산 방법과 비교하거나 시행착오를 통해. 반면에 재료 데이터, 벤치마크 및 모델, 공개적으로 사용 가능한 교육이 부족하다는 점이 부각되고 있습니다 데이터를 수집하고 사전 학습된 모델을 엽니다. 이 문제를 해결하기 위해 메타 페어를 개최합니다 오픈 머티리얼즈 2024(OMAT24) 대규모 오픈 데이터 세트 및 사전 학습된 모델 세트가 함께 제공됩니다. OMat24에는 1억 1천만 개 이상의 모델이 포함되어 있습니다 밀도 함수 이론(DFT) 계산은 구조 및 구성 다양성. EquiformerV2 모델은 최첨단 성능을 달성합니다 매트벤치 디스커버리 리더보드의 성능 및 예측 기능 0.9 이상의 F1 점수와 각각 20meV/원자의 정확도. 모델 크기의 영향을 살펴봅니다, 보조 노이즈 제거 목표 및 범위 전반의 성능 미세 조정 OMat24, MPtraj, 알렉산드리아를 포함한 데이터 세트의 공개 릴리스 OMat24 데이터 세트 및 모델을 통해 연구 커뮤니티는 AI 지원 재료 과학 분야에서 더욱 발전하기 위한 노력을 기울이고 있습니다.
URL: https://huggingface.co/papers/2410.12771

제목: MuVi: 시맨틱 정렬과 리듬적 동기화를 통한 비디오-음악 세대
번역된 요약:
동영상의 시각적 콘텐츠에 맞는 음악을 생성하는 것은 다음과 같습니다 시각적 의미에 대한 깊은 이해가 필요하기 때문에 어려운 작업입니다 멜로디, 리듬, 역학이 다음과 조화를 이루는 음악을 생성하는 것을 포함합니다 시각적 내러티브. 본 논문은 효과적으로 새로운 프레임워크인 MuVi를 제시한다 이러한 과제를 해결하여 다음과 같은 응집력과 몰입형 경험을 향상시킵니다 시청각 콘텐츠. MuVi는 특별히 설계된 비디오 콘텐츠를 통해 분석합니다 시각적 어댑터를 사용하여 상황 및 시간적으로 관련된 기능을 추출합니다. 다음 사항 기능은 비디오의 분위기뿐만 아니라 음악을 생성하는 데 사용됩니다 테마뿐만 아니라 리듬과 속도도 중요합니다. 또한 대조적인 요소도 소개합니다 동기화를 보장하는 음악 시각 사전 훈련 계획은 다음과 같습니다 음악 문구의 주기성 특성. 또한, 우리는 플로우 matching 기반 음악 생성기는 context 내 학습 능력을 갖추고 있어 다음을 가능하게 합니다 생성된 음악의 스타일과 장르를 제어합니다. 실험 결과 MuVi는 오디오 품질과 시간적 동기화. 생성된 뮤직 비디오 샘플은 다음에서 확인할 수 있습니다 https://muvi-v2m.github.io .
URL: https://huggingface.co/papers/2410.12957

제목: LLM은 정치적 올바름을 가지고 있나요? AI 시스템의 윤리적 편견과 탈옥 취약점 분석
번역된 요약:
대형 언어 모델(LLM)은 다음과 같은 분야에서 인상적인 숙련도를 보여줍니다 다양한 작업은 'jail 파손'과 같은 잠재적 안전 위험을 제시하며, 다음과 같은 경우 악의적인 입력은 LLM에게 유해한 콘텐츠를 생성하도록 강요할 수 있습니다. 주소 지정 이러한 문제로 인해 많은 LLM 개발자는 다음과 같은 다양한 안전 조치를 시행했습니다 이러한 모델을 정렬합니다. 이 정렬에는 데이터를 포함한 여러 기술이 포함됩니다 사전 교육, 감독된 미세 조정, 강화 학습 중 필터링 사람의 피드백과 레드 티밍 운동에서. 이러한 방법은 종종 다음과 같은 이점을 제공합니다 정치적 올바름(PC)과 유사한 의도적이고 의도적인 편견 LLM의 윤리적 행동을 보장합니다. 본 논문에서는 안전 목적으로 LLM에 주입된 의도적 편향성 및 방법 검토 이러한 안전 정렬 기술을 우회합니다. 특히 이러한 의도적인 편견으로 인해 GPT-4o 모델의 탈옥 성공률은 다음과 같습니다 비binary 및 시스젠더 키워드 간에는 20%, 백인과 시스젠더 간에는 16%의 차이가 있습니다 프롬프트의 다른 부분이 동일한 경우에도 검은색 키워드를 사용합니다. 우리는 PCJailbreak의 개념을 소개하고 다음에 따른 내재적 위험을 강조합니다 이러한 안전으로 인한 편견. 또한 효율적인 방어를 제안합니다 방어력을 주입하여 탈옥 시도를 방지하는 방법 PCDefense 생성 전 프롬프트. PCDefense는 다음과 같은 매력적인 대안으로 사용됩니다 라마-가드와 같은 가드 모델은 다음과 같은 추가 추론 비용이 필요합니다 텍스트 생성. 우리의 연구 결과는 LLM 개발자가 다음을 수행해야 할 긴급한 필요성을 강조합니다 안전을 설계하고 구현할 때 보다 책임감 있는 접근 방식을 채택합니다 방안.
URL: https://huggingface.co/papers/2410.13334

제목: LoLDU: 매개변수 효율적인 미세 조정을 위해 하위 다이어그램-상위 분해를 통한 낮은 순위 적응
번역된 요약:
모델 규모의 급속한 성장으로 인해 상당한 계산이 필요해졌습니다 미세 조정을 위한 리소스. 로우 랭크 적응(LoRA)과 같은 기존 접근 방식 에서 대규모 업데이트된 매개 변수를 처리하는 문제를 해결하려고 했습니다 전체 미세 조정. 그러나 LoRA는 무작위 초기화 및 최적화를 활용합니다 업데이트된 가중치를 근사화하기 위해 낮은 순위 행렬을 사용하면 다음과 같은 결과가 발생할 수 있습니다 최적이 아닌 수렴 및 전체 미세 조정과 비교했을 때 정확도 갭이 있습니다. To 이러한 문제를 해결하기 위해 매개변수 효율적인 미세 조정인 LoLDU를 제안합니다 학습 가능한 매개변수를 2600배 크게 줄이는 (PEFT) 접근 방식 일반 PEFT 방식과 비교하면서도 비슷한 성능을 유지합니다. LoLDU는 하위 진단-상위 분해(LDU)를 활용하여 하위 순위를 초기화합니다 더 빠른 수렴과 직교성을 위한 행렬. 우리는 변환을 확장하기 위한 대각선 행렬. 우리가 아는 한, LoLDU는 모든 PEFT 접근 방식 중 매개 변수가 가장 적습니다 4개의 명령어 following 데이터 세트, 6개의 내추럴에 걸친 광범위한 실험 언어 이해(NLU) 데이터 세트, 8개의 이미지 분류 데이터 세트 및 여러 모델 유형(LLaMA2, RoBERTa, ViT)이 포함된 이미지 생성 데이터 세트 및 안정적 확산), 포괄적이고 상세한 분석 제공. 우리의 오픈 소스 코드는 다음 위치에서 액세스할 수 있습니다 https://github.com/SKDDJ/LoLDU {https://github.com/SKDDJ/LoLDU }.
URL: https://huggingface.co/papers/2410.13618

제목: AERO: 효율적인 비공개 추론을 위한 소프트맥스 전용 LLM
번역된 요약:
독점 언어 모델의 보편성으로 인해 개인정보 보호에 대한 우려가 커지고 있습니다 비공개 추론(PI)의 필요성을 강조하는 사용자의 민감한 데이터의 경우, 여기서 추론은 암호화된 입력에 대해 직접 수행됩니다. 그러나 현재 PI 방법은 엄청나게 높은 통신 및 지연 시간 오버헤드에 직면해 있습니다, 주로 비선형 연산으로 인해 발생합니다. 이 논문에서는 비선형성의 역할을 이해하기 위한 종합적인 분석 트랜스포머 기반 디코더 전용 언어 모델. 4단계인 AERO를 소개합니다 기존 LLM 아키텍처를 개선하는 아키텍처 최적화 프레임워크 레이어노름과 같은 비선형성을 체계적으로 제거하여 효율적인 PI를 구현합니다 그리고 GELU와 FLOP 수 감소. 우리는 처음으로 효율성을 위해 맞춤화된 FLOP가 훨씬 적은 소프트맥스 전용 아키텍처 PI. 또한 개선하기 위해 새로운 엔트로피 정규화 기법을 고안합니다 소프트맥스 전용 모델의 성능. AERO는 최대 4.23배 달성 커뮤니케이션 및 1.94배 지연 시간 단축. 우리는 그 효과를 검증합니다 AERO를 최신 기술과 비교하여 벤치마킹합니다.
URL: https://huggingface.co/papers/2410.13060

제목: 조건 대조 정렬을 통한 안내 없는 AR 시각적 생성을 위한 방법
번역된 요약:
분류기 없는 지침(CFG)은 다음을 개선하기 위한 중요한 기술입니다 시각적 생성 모델의 샘플 품질. 그러나 자동 회귀(AR)에서는 멀티모달 세대, CFG는 언어 간의 설계 불일치를 도입합니다 시각적 콘텐츠로, 서로를 통합하려는 디자인 철학과 모순됩니다 시각적 AR을 위한 양식. 언어 모델 정렬 방법에서 영감을 받아 우리는 다음을 촉진하기 위한 조건 대조 정렬(CCA) 제안 고성능의 안내 없는 AR 시각적 생성 및 분석 가이드 샘플링 방법과의 이론적 연관성. 가이드 방법과 달리 이상적인 샘플링 분포인 CCA를 달성하기 위해 샘플링 프로세스를 변경합니다 동일한 배포 대상에 맞도록 사전 학습된 모델을 직접 미세 tunes합니다. 실험 결과, CCA가 지침 없는 기능을 크게 향상시킬 수 있음을 보여줍니다 단 한 번의 미세 조정으로 테스트한 모든 모델의 성능(sim 1₩%) 사전 훈련 에포크)의 사전 훈련 데이터 세트는 가이드 샘플링과 동등한 수준입니다 방법. 이를 통해 AR 비주얼에서 가이드 샘플링의 필요성을 크게 제거할 수 있습니다 생성 및 샘플링 비용을 절반으로 줄였습니다. 또한 교육을 조정하여 매개변수, CCA는 표본 다양성과 충실도 사이의 균형을 달성할 수 있습니다 CFG와 유사합니다. 이는 강력한 이론적 연관성을 실험적으로 확인합니다 언어-targeted 정렬과 시각-targeted 안내 방법 사이, 이전에 독립적이었던 두 가지 연구 분야를 통합합니다. 코드 및 모델 가중치: https://github.com/thu-ml/CCA .
URL: https://huggingface.co/papers/2410.09347

제목: TransAgent: 이기종 에이전트 협업을 통한 비전-언어 기반 모델 전송
번역된 요약:
최근 비전 언어 기반 모델(예: CLIP)이 다음과 같은 기능을 선보이고 있습니다 대규모 이미지 텍스트 사전 학습으로 인한 전이 학습의 힘.
그러나 다운스트림 작업의 대상 도메인 데이터는 매우 다를 수 있습니다 사전 훈련 단계부터, 단일 모델로는 다음과 같은 작업을 수행하기 어렵습니다 일반화를 잘합니다. 또는 다음과 같은 다양한 전문가 모델이 존재합니다 다양한 비전 및/또는 언어 지식을 사전에 학습할 수 있습니다 양식, 작업, 네트워크 및 데이터 세트. 안타깝게도 이러한 모델은 다음과 같습니다 이기종 구조를 가진 "isolated 에이전트" 및 이를 통합하는 방법
CLIP과 유사한 모델을 일반화하기 위한 지식은 아직 완전히 탐구되지 않았습니다. To 이러한 격차를 해소하기 위해 일반적이고 간결한 TransAgent 프레임워크를 제안합니다 고립된 에이전트의 지식을 통합된 방식으로 전송합니다 는 클립이 다중 소스 지식 증류로 일반화되도록 효과적으로 안내합니다. 이러한 독특한 프레임워크를 통해 11개의 이기종과 유연하게 협업합니다 에이전트가 추가 비용 없이 비전 언어 기반 모델에 힘을 실어줍니다 추론 단계. 마지막으로 TransAgent는 최첨단 기술을 달성합니다 11개의 시각 인식 데이터 세트에 대한 성능. 동일한 로우 샷 설정에서, 평균 약 10%, 유로SAT에서 20%로 인기 있는 CoOp를 능가합니다 대규모 도메인 이동이 포함되어 있습니다.
URL: https://huggingface.co/papers/2410.12183

제목: SBI-RAG: 스키마 기반 교육 및 검색 증강 생성을 통해 학생들의 수학 단어 문제 해결력 향상
번역된 요약:
많은 학생들이 수학 단어 문제(MWP)로 어려움을 겪고 있으며, 종종 이를 발견합니다 주요 정보를 식별하고 적절한 수학적 정보를 선택하기 어렵습니다 operations.Schema 기반 명령(SBI)은 다음과 같은 증거 기반 전략입니다 학생들이 구조에 따라 문제를 분류하고 개선하는 데 도움이 됩니다 문제 해결 정확도. 이를 기반으로 우리는 스키마 기반을 제안합니다 다음과 같은 명령어 검색-증강 생성(SBI-RAG) 프레임워크 는 대규모 언어 모델(LLM)을 통합합니다.우리의 접근 방식은 단계별로 강조합니다 스키마를 활용하여 솔루션 생성을 유도하여 추론합니다. 우리는 그것을 평가합니다 GSM8K 데이터 세트의 성능, GPT-4 및 GPT-3.5 터보와 비교, 솔루션 품질을 평가하기 위해 "추론 점수" 지표를 도입합니다. 우리의 연구 결과에 따르면 SBI-RAG는 추론 명확성과 문제 해결을 향상시킵니다 정확성, 잠재적으로 학생들에게 교육적 혜택 제공
URL: https://huggingface.co/papers/2410.13293

제목: 고품질 데이터를 핵심으로 하는 LLM에서 긴 출력을 잠금 해제하는 최소 튜닝
번역된 요약:
대형 언어 모델이 더 긴 컨텍스트를 지원하기 위해 빠르게 진화함에 따라 더 긴 길이로 출력을 생성하는 능력의 현저한 차이입니다. 최근 연구에 따르면 이러한 불균형의 주요 원인은 다음에서 발생할 수 있습니다 정렬 훈련 중에 긴 출력을 가진 데이터의 부족. 이에 비추어 볼 때 관찰, 다음과 같은 데이터로 기초 모델을 재 align하려고 시도합니다 공백을 메우고 다음과 같은 경우 긴 출력을 생성할 수 있는 모델을 생성합니다 지시했습니다. 본 논문에서는 데이터 품질이 튜닝에 미치는 영향을 살펴봅니다 긴 출력에 대한 모델과 시작점에서 그렇게 할 수 있는 가능성 인간이 정렬한 (명령 또는 채팅) 모델의 경우. 신중한 데이터 큐레이션을 통해 다음을 확인할 수 있습니다 튜닝한 제품에서도 유사한 성능 향상을 달성할 수 있습니다 학습 데이터 인스턴스와 컴퓨팅의 극히 일부만 포함된 모델. In 또한, 우리는 다음을 적용하여 이러한 접근 방식의 일반화 가능성을 평가합니다 여러 모델로 recipes을 튜닝합니다. 우리의 연구 결과에 따르면 용량은 다음과 같습니다 긴 출력을 생성하는 데는 기본적으로 여러 모델에 따라 다릅니다 라이트 컴퓨팅을 사용하여 고품질 데이터로 일관되게 조정하는 접근 방식 실험한 모든 모델에서 눈에 띄는 개선 효과를 얻었습니다 롱라이팅 기능 튜닝을 위한 큐레이팅된 데이터 세트인 모델 튜닝 및 평가와 미세 tuned의 구현 공개적으로 accessed할 수 있는 모델입니다.
URL: https://huggingface.co/papers/2410.10210