

코멘트 생성 품질 Human evaluation 가이드라인

부스트캠프 AI Tech NLP-04

평가 개요

모델이 사용자의 일기에 대해 생성한 코멘트에 대한 품질을 정성적으로 평가한다. 좋은 코멘트란, ① 문장이 문법적·의미적·논리적으로 자연스럽고, ② 일기에서 언급하는 사건에 대한 구체적인 언급을 해주며 ③ 일기에서 드러나는 감정에 대해 공감해주는 코멘트라고 할 수 있다.

이러한 특성을 평가할 수 있는 방법으로 Meena([Adiwardana et al., 2020](#))에서 제안한 SSA (Sensibleness and Specificity Average)를 활용할 수 있다. SSA 란 합리성(Sensibleness)와 구체성(Specificity)의 평균 점수로서, 챗봇의 발화가 얼마나 사람 같은지를 측정하는 지표로 제안되었다. 이와 더불어 코멘트가 일기에 대해 얼마나 공감적인지를 측정하는 '공감성'이라는 지표도 활용한다.

합리성(Sensibleness)

기준

- 주어진 대화 맥락 내에서 말이 되는 발화인가?
- 일반적인 상식, 사실에 부합하는가?
- 문장이 문법적·의미적으로 자연스러운가?

예시

일기	
우리 교회는 3층에 있다. 엄마가 갑자기 다리가 아프다면서 내 머리를 눌렀다. 내가 작아지는 기분이었다.	
합리성을 충족하는 코멘트 예시	합리성을 충족하지 않는 코멘트 예시
엄마가 다리가 아프다니 안타깝군요. 그래도 기도를 하고 함께 응원해주면 엄마가 빨리 나아질 거예요.	좋은 생각이네요! 엄마 다리가 아파서 더 작아지셨을 것 같아요.

'작아지는'의 주체는 '나'이므로, '엄마'가 작아진다는 언급은 주어진 대화 맥락 내에서 말이 되지 않음.

일기
해피(히야신스)가 자랐다. 깜짝 놀랐다. 히야신스는 성장이 너무 빠르다. 사랑을 듬뿍줘야지 잘 자랄 것이다.

합리성을 충족하는 코멘트 예시	합리성을 충족하지 않는 코멘트 예시
오 너무 신기하네요! 당신이 해주는 사랑과 관심이 히야신스의 빠른 성장에 큰 도움이 되었을 것 같아요! 계속 응원할게요!	좋은 일기예요! 게 되는 것 같아 다행이네요!

'게 되는 것 같아 다행이네요'.
문장이 비문법적임.

일기
ㅠㅠ 너무 짜증나... 흑.. 흑흑.. 나도 내 시간이 있는데 약속을 안 지키는 친구가 있어서 짜증이 난다,,,

합리성을 충족하는 코멘트 예시	합리성을 충족하지 않는 코멘트 예시
앗, 짜증이 나는 상황이군요. 그래도 이런 상황에도 부담가지지 말고 힘내세요. 약속을 지키는 것이 중요하다고 생각해요. 힘내세요!	정말 좋은 생각이네요! 당신의 약속을 지키지 않는 것이 중요해요. 그리고 다른 사람과 함께 즐거운 시간을 보내세요!

일반적인 상식과 부합하지 않음

평가 방법

- 합리성을 만족한다면 1, 그렇지 않다면 0을 부여한다.

구체성(Specificity)

기준

- 주어진 맥락(일기)에 대해 구체적으로 언급하는가? 혹은 주어진 맥락과 밀접하게 관련된 언급을 하고 있는가?

예시

일기
카레를 먹었는데 너무 너무 카레가 맛있었다. 그리고 나서 아빠와 함께 축구를 했다. 내가 10:3으로 이겨서 기분이 좋았다. 저녁에는 5-1 책들도 샀다. 오늘 저녁은 콩나물국밥을 먹었더니 너무 맛있어서 좋았다.

구체성을 충족하는 코멘트 예시	구체성을 충족하지 않는 코멘트 예시
좋은 하루 보냈네요! 5-1 책들도 샀다는 것이 너무 좋네요. 카레라는 것도 맛있었고, 축구도 잘하셨어요. 내일도 화이팅하세요!	정말 좋은 기억이 될 것 같아요! 내일도 즐거운 시간 보내세요!

주어진 맥락에서만 가능한
코멘트가 아니라, 어떤 일기에
대해서도 가능한 코멘트임

일기
심폐소생술 수업을 했다. 흉부압박은 필수다. 그런데 인공호흡은 조금 꺼림칙하다. 선생님이 인공 호흡은 꺼림칙 하면 꼭 해야 되는게 아니라고 하셨다. 다행이다. 심폐소생술 기계로 심폐소생술을 해보았는데 신기했다.

구체성을 충족하는 코멘트 예시	구체성을 충족하지 않는 코멘트 예시
인공호흡이 꺼림칙하더라도 흉부 압박이 꼭 필요한 것은 분명합니다. 심폐소생술 기계를 사용해본 경험은 소중한 경험인 것 같아요. 이런 경험을 통해 더 많은 사람들을 구할 수 있는 능력을 갖추 수 있을 테니까요	정말 다행이네요! 정말 좋은 일기에요!

주어진 맥락에서만 가능한
코멘트가 아니라, 어떤 일기에
대해서도 가능한 코멘트임

평가 방법

- 구체성을 만족한다면 1, 그렇지 않다면 0을 부여한다.

공감성(Sympathy)

기준

- 코멘트가 일기에서 드러나는 감정에 대해 잘 이해하고 공감해주거나, 혹은 위로해주고 있는가?
 - 감정에 대해 잘 이해하고 공감해주어야 한다. 누군가를 위로할 때 많이 쓰는 관용적인 표현을 썼더라도, 억지로 위로해준다는 느낌이 들거나, 혹은 일기의 감정을 전혀 이해하지 못한 채 위로만 해주려는 느낌을 받는다면 '공감성'이 없는 것으로 판단한다.

예시

일기	
육상 경기를 보러갔다. 길을 잃어서 못볼줄 알았는데 길을 잘 아는 사람들을 만나서 다행이었다. 400m를 달리는 남자 고등학생 기록이 45.12초였다. 나도 그렇게 빨리달리고 싶고 앞으로 달리기연습을 해야겠다	
공감성을 충족하는 코멘트 예시	공감성을 충족하지 않는 코멘트 예시
좋은 일이었군요! 길을 잘 아는 사람들을 만나서 다행이었겠네요. 달리기 연습을 열심히 해서 좋은 결과를 얻으셨으면 좋겠어요!	육상 경기를 보러갔군요. 길을 잘 아는 사람을 많났군요. 앞으로 달리기연습을 해야겠네요.

일기에서 드러나는
감정에 대한 언급이 없음

일기		
“만약에”라는 한글자의 단어는 터무니없는 희망을 지니게 하기도 한다. 만약이라는건 정말 슬픈 단어다. 만약에 내가 로또에 당첨되면 집을 사야지, 차를 사야지, 명품을 사야지라는 터무니없는 희망.		
공감성을 충족하는 코멘트 예시	공감성을 충족하지 않는 코멘트 예시	
'만약'이라는 단어는 참 슬픈 단어인 것 같아요. 오늘은 이것저것 생각이 많은 하루를 보내셨나 봐요. 언젠가 원하는 모든 게 이뤄지기를 바랍니다.	만약이라는 단어는 우리가 꿈꾸고, 희망하는 것을 의미하는 것 같아요. 하지만 만약에 우리가 꿈과 희망을 잃게 된다면, 어떻게 해야할까요? 그러면 우리는 더 열심히 해야하고 더 노력해야하는 것 같아요.	좋은 생각이네요! '만약에'라는 단어는 우리가 꿈을 꾸는 데 도움이 되는 단어들이라는 것을 잊지 마세요. 당신의 꿈을 이루기 위해 노력해보는 건 어떨까요? 그리고 터무니없는 희망을 가지게 되는 것도 좋은 것 같아요.

일기의 주된 감정인
슬픈 감정에 대해
위로해주는 것이 아님

일기의 말에 동의해주고
있기는 하지만, 일기의 주된
감정인 슬픈 감정에 대해
위로해주는 것이 아님

평가 방법

- 공감성을 만족한다면 1, 그렇지 않다면 0을 부여한다.