AI 활용 수업의 실제와 적용

인공지능에게 답을 맡기지 않는 수업 설계

인공지능 스무고개 — 빅데이터의 정체를 찾아라

수행평가 설계서 · 평가요소·평가방법·평가기준 포함
고등학교 「빅데이터의 활용」·「통합과학」 수행평가 자료
대상: 고등학교 1학년

지도교사: (최 성 훈)

숭의여자고등학교

2026. 9.

목 차

Ⅰ. 문서 개요 — 학습과 탐구의 관점에서 본 인공지능

Ⅱ. 제1부 수행평가 — 인공지능 스무고개: 빅데이터의 정체를 찾아라

Ⅴ. 결론

Ⅵ. 운영상 유의점과 한계

부록 A. 제1부 진행 대본 — 스무고개 힌트 시퀀스 예시

I

문서 개요

1. 학습과 탐구의 관점에서 본 인공지능

인공지능이 답을 대신하는 시대, 교육은 무엇을 잃고 있는가

챗GPT를 비롯한 생성형 인공지능이 대중화되면서, 우리는 궁금한 것을 검색창에 입력하던 시대에서 인공지능에게 바로 물어보는 시대로 넘어왔다. 복잡한 수학 문제부터 논술형 에세이까지, 인공지능은 몇 초 만에 그럴듯한 답을 내놓는다. 편리함은 분명하다. 하지만 이 편리함이 교육 현장, 특히 중고등학교 교실에서는 심각한 문제를 낳고 있다.

과제가 '나의 것'이 아니게 되는 순간

가장 먼저 드러난 문제는 과제 제출 방식의 붕괴다. 많은 중고등학생들이 독서감상문, 영어 에세이, 심지어 수학 서술형 문제까지 인공지능에게 맡기고 있다. 문제는 그 결과물이 겉으로는 흠잡을 데 없이 매끄럽다는 점이다. 문법 오류도 없고, 논리 전개도 자연스럽다. 교사가 한눈에 인공지능이 작성했다고 판별하기 어려운 경우도 많다.

이렇게 제출된 과제는 학생의 사고 과정을 전혀 반영하지 않는다. 원래 과제란 학생이 스스로 자료를 찾고, 생각을 정리하고, 시행착오를 겪으며 표현력을 기르는 과정이어야 한다. 그런데 인공지능이 그 과정을 통째로 대신해버리면, 학생에게 남는 것은 '제출 완료'라는 결과 뿐, 배움은 사라진다. 더 심각한 것은 이런 방식에 익숙해진 학생일수록 정작 시험이나 실제 발표 상황에서 자신의 생각을 표현하지 못한다는 점이다. 과제와 실력 사이의 괴리가 점점 벌어지고 있는 것이다.

평가는 왜 무력해지는가

교사 입장에서도 곤란하기는 마찬가지다. 과제를 통한 평가가 더 이상 학생의 실제 이해도를 반영하지 못하기 때문이다. 표절 검사 도구로는 인공지능이 생성한 글을 걸러내기 어렵고, 설령 의심이 가더라도 이를 명확히 증명하기는 쉽지 않다. 결국 많은 학교가 과제의 비중을 줄이거나 아예 손으로 쓰는 시험으로 회귀하는 임시방편을 택하고 있지만, 이는 근본적인 해결책이 되지 못한다. 과제라는 학습 방식 자체가 신뢰를 잃어가고 있는 셈이다.

그래서 필요한 것: 인공지능이 대신 풀 수 없는 문제

이 문제의 핵심은 결국 '질문의 설계'에 있다. 인공지능이 즉답을 내놓을 수 있는 것은, 기존의 과제 형식이 이미 어딘가에 정답이 존재하는 지식 재현형 질문이기 때문이다. "OO 소설의 주제를 설명하시오", "다음 영어 지문을 요약하시오" 같은 질문은 인공지능이 가장 잘하는 영역이다.

따라서 지금 교육에 필요한 것은 인공지능을 막는 것이 아니라, 인공지능이 대신 답할 수 없는 문제를 개발하는 일이다. 예를 들면 다음과 같은 방향이다.

- **개인의 경험과 맥락을 반드시 요구하는 질문**: 자신이 겪은 구체적 사건, 우리 학교·우리 동

네의 사례를 근거로 답하게 하여 일반화된 답변을 무력화한다.

- **과정을 기록하게 하는 과제**: 최종 결과물이 아니라 생각이 바뀌어 간 과정, 초안과 수정본의 변화를 함께 제출하게 한다.
- **실시간·대면 방식의 평가**: 짧은 구술 발표나 즉석 질의응답을 병행해, 글의 내용을 스스로 설명할 수 있는지 확인한다.
- **인공지능을 도구로 인정하고 활용법 자체를 평가하는 방식**: 인공지능의 답을 비판적으로 검토하고 수정·반박하게 하여, 오히려 비판적 사고력을 기르는 과제로 전환한다.

인공지능은 이제 되돌릴 수 없는 현실이다. 교육이 해야 할 일은 인공지능과 경쟁하는 것이 아니라, 인공지능이 대신할 수 없는 사고의 영역을 다시 설계하는 것이다. 그럴 때 비로소 과제는 다시 '학생 자신의 것'이 될 수 있을 것이다.

2. 배경 — 왜 이런 수업이 필요했는가?

계기는 평가 설계 회의가 아니라 실제 수업 현장에서 발생했다. 학생들과 과학탐구실험 프로젝트를 진행하던 중 문제가 발생했고, 교사가 함께 해결에 나섰다. 그런데 학생들은 "인공지능에게 물어보면 된다"며 먼저 인공지능에 질문을 넣었다. 결과를 확인해 보니 두 가지 문제가 있었다 — 학생들이 넣은 질문 자체가 상황을 정확히 설명하지 못했고, 인공지능이 내놓은 답도 일반적인 수준의 기술이라 실제 문제에는 그대로 적용되지 않았다. 교사가 이 점을 짚어 주었음에도 학생들은 교사의 설명보다 인공지능의 답을 더 신뢰하려는 태도를 보였다.

이러한 모습에서 학생들에게 인공지능은 이미 "물어보면 뭐든 답해주는 절대적인 존재"로 굳어졌다는 것이다. 스마트폰으로 궁금한 걸 검색하던 습관이 어느새 인공지능에게 묻는 습관으로 자연스럽게 옮겨갔고, 학생들은 그 답을 의심 없이 받아들이는 데 익숙해졌다. 숙제를 하다 막히면 참고서를 펼치기보다 인공지능 창을 여는 것이 더 빠르고 당연한 선택이 된 것이다.

문제는 정작 인공지능에게 상황을 제대로 설명하고, 그 답이 맞는지 스스로 따져보는 힘은 아직 없다는 것이다. 많은 학생들이 질문을 던질 때 자신이 처한 맥락이나 조건을 충분히 담지 못한 채 짧고 막연하게 묻고, 돌아온 답을 그대로 베껴 쓴다. 그 답이 실제로 자신의 상황에 맞는지, 틀린 부분은 없는지 되짚어보는 과정은 생략된다. 결국 인공지능을 "사용"은 하지만 "다루지"는 못하는 셈이다.

이 문서는 인공지능을 못 쓰게 막으려는 게 아니라, 그것을 어떻게 대하고 다뤄야 하는지를 학생들이 직접 부딪혀보며 익히게 하려는 것이 이 평가의 목표다. 질문을 정확하게 던지는 법, 답을 곧이곧대로 믿지 않고 검증하는 법, 그리고 그 과정에서 스스로 생각하는 힘을 잃지 않는 법을 몸으로 익히도록 설계한 것이다. 결국 이 평가가 겨냥하는 것은 인공지능과의 경쟁이 아니라, 인공지능과 함께 있어도 흔들리지 않는 사고력을 기르는 일이다.

Ⅱ

인공지능 스무고개: 빅데이터의 정체를 찾아라

1. 수업 개요

구분	내용
교과·과목	고등학교「빅데이터의 활용」
대상·형태	수강 학급 전체, 4인 1팀 편성
차시	1차시(수업시간 1시간) 내 완결
평가유형	수행평가(팀 산출물 + 시간 준수)
도구	학생이 접근 가능한 인공지능 서비스 전체(종류 제한 없음)
산출물	숫자 데이터의 의미 서술 + 그래프

2. 발단— 과학탐구 프로젝트에서 벌어진 일

학생들과 과학탐구실험 프로젝트를 진행하던 중 발생한 문제를 교사가 함께 해결하려 했으나, 학생들은 먼저 인공지능에게 물어보고 그 답만 믿으려 했다. 그러나 학생들이 인공지능에 입력한 질문은 상황을 정확히 설명한 것이 아니었고, 인공지능의 답 역시 일반적인 기술 수준이라 실제로는 적용되지 않았다. 이 경험이 수행평가 설계의 직접적인 계기가 되었다. 인공지능이 절대적인 만능이 아니라는 것을 확인시킬 필요가 있었다.

3. 설계 의도

- 인공지능이 한 번에 정답을 내어 주기 어려운 문제를 의도적으로 구성.
- 정답을 바로 알려주지 않고 "예/아니오" 힌트로 범위를 좁혀 가는 스무고개 구조를 사용하여 학생이 인공지능에게 던지는 질문 자체를 스스로 다시 만들어 가도록 유도.
- 팀 단위로 여러 인공지능을 동시에 활용하게 하여, 서비스별 성능 차이(그래프 작성 가능 여부 등)를 직접 비교·체감하게 함.

4. 문제 자료 구성

케플러의 제3법칙(조화의 법칙) 수치 데이터 하나만으로 문제를 구성하면 인공지능이 지나치게 빨리 정답을 찾아낼 위험이 있었다. 그래서 성격이 전혀 다른 세 가지 수치 자료를 함께 제시해,

항목명과 단위를 모두 지운 표 형태로 섞어 배부했다.

자료	출처·특징	쉽게 풀리지 않는 이유
① 케플러 제3법칙 (조화의 법칙) 수치	티코 브라헤의 관측 기록을 케플러가 10년 넘게 분석·계산해 얻은 행성 공전주기·궤도 장반경 수치	원자료에 있는 컬럼에 들어있는 값을 제거하고 숫자 형태의 데이터만 제공
② 쌍성의 절대등급 질량과 광도 변화 수치	지구과학 교과서에 수록된, 시간에 따른 쌍성 질량과 밝기(등급) 변화 자료	등급 체계 자체가 "값이 작을수록 밝다"는 비직관적 구조라, 인공지능도 축의 방향을 해석하는 단계에서 혼동을 일으킨다
③ 학생 실측 심장박동 신호 데이터	학생들이 실험실에서 직접 측정해 엑셀로 정리한 실험 데이터	실험실에서 나온 값이라는 것 외에는 아무 맥락 정보가 없어, 공학 데이터나 화학 데이터로 오인되기 쉽다

** 세 자료는 모두 항목명·단위·출처를 지운 순수 숫자 표 형태로만 제시했다.*

5. 수업 방법 및 진행 절차

시간	활동	교사 역할
0~5분	항목명이 지워진 숫자 표 3종 배부, 4인 1팀 확인, 규칙 안내(모든 인공지능 사용 가능, 50분 내 정답 제출)	규칙 공지, 팀 편성 확인
5~30분	팀별로 접속 가능한 인공지능에 자료를 입력해 정체를 추정하고, 필요하면 다른 인공지능에도 같은 자료를 넣어 답을 비교	정답을 알려주지 않고, 팀이 질문할 때만 "예/아니오"로 스무고개 힌트 제공 (부록 A 참조)
30~40분	찾아낸 수치의 의미를 서술하고, 인공지능으로 그래프를 작성	그래프 작성 가능 여부를 확인하며 순회, 시간 마감 예고
40~50분	팀별 결과(의미 서술 + 그래프) 제출, 여건이 되면 사용한 인공지능 비교 소감 간단 공유	제출 확인 및 마감 시각 기준 완료 여부 판정

6. 운영 규칙

(1) 계정·사용량 관리

팀당 50분이라는 제한 시간 안에 여러 차례 인공지능에 질문을 반복해야 하므로, 한 학생의 무료 사용량이 소진되면 다른 팀원의 계정으로 로그인해 이어가는 방식으로 3~4개 계정을 순환 사용하도록 사전에 안내한다. 이를 통해 팀 전체가 시간 내에 사용량 제한 없이 과제를 이어갈 수 있었다.

(2) 채점 규칙

실제 수업에서 적용한 규칙은 단순했다 — 50분 내에 정답(수치의 의미 서술 + 그래프)을 제출하면 만점, 시간 내에 완료하지 못하면 1점을 감점한다. 아래 7~9절의 평가 요소·방법·기준은 이 "완료"를 판단하는 기준을 구체화하고, 교사의 순회 관찰을 채점에 반영할 수 있도록 정리한 것이다.

7. 평가 요소

요소	설명
① 데이터 정체 규명의 정확성	세 자료 각각이 무엇을 나타내는 값인지 정확히 서술했는가
② 그래프 작성의 타당성	값의 성격에 맞는 그래프 형태(꺾은선·산점도 등)를 선택하고 축·단위를 바르게 표기했는가
③ 인공지능 활용 전략	같은 자료를 여러 인공지능에 입력해 답을 비교했는가, 힌트를 반영해 질문을 스스로 좁혀 나갔는가
④ 시간 내 완료와 협업	50분 내 팀 전체 결과를 제출했는가, 계정 순환 등 팀 내 역할 분담이 이루어졌는가

8. 평가 방법

- 팀 산출물 평가: 제출된 정답지(수치의 의미 서술)와 그래프를 확인
- 수행 관찰: 교사가 순회하며 인공지능 활용 방식(질문을 바꾸는지, 여러 서비스를 비교하는지)을 체크리스트로 관찰
- 시간 준수 확인: 50분 마감 시점 기준 제출 여부 확인

9. 평가 기준

요소	상	중	하
데이터 정체 규명	세 자료 모두 정확히 규명	두 자료만 정확히 규명	한 자료 이하 규명, 또는 시간 내 미제출
그래프 작성	세 자료 모두 값 성격에 맞는 그래프와 정확한 축·단위 표기	그래프는 작성했으나 일부 형식이나 표기가 미흡	그래프 미작성, 또는 값 성격과 무관한 그래프
인공지능 활용 전략	2개 이상 인공지능을 비교하고, 힌트를 반영해 질문을 스스로 재구성	인공지능은 여러 개 사용했으나 질문 재구성은 미흡	처음 얻은 답을 그대로 채택
시간·협업	50분 내 전원이 참여해 제출	시간 내 제출했으나 팀원 간 참여가 불균등	50분 초과(감점 규칙 적용)

** 실제 수업에서는 "시간 내 완료 시 만점 / 미완료 시 1점 감점"이라는 규칙으로 채점했으며, 위 표는 그 "완료"의 질을 판단하기 위한 참고 기준이다.*

10. 수업에서 확인된 것

계획하지 않았던 부가적 학습 효과도 여럿 있었다.

먼저, 어떤 인공지능은 그래프를 그릴 수 있고 어떤 인공지능은 그러지 못한다는 사실을 학생들이 직접 부딪혀보며 확인했다. 같은 질문을 던져도 어떤 서비스는 데이터를 곧바로 시각화해주는 반면, 어떤 서비스는 말로만 설명하고 끝나버렸다. 이 차이를 몸으로 겪으면서 학생들은 "인공지능은 다 똑같이 잘한다"는 막연한 생각에서 벗어나, 도구마다 잘하는 영역과 못하는 영역이 따로 있다는 것을 자연스럽게 체감했다.

또한 이전에는 구글 제미나이·챗지피티 정도만 알고 있던 학생들이, 그록·퍼플렉시티·클로드 등 훨씬 다양한 인공지능 서비스가 존재한다는 사실을 새롭게 알게 되었다. 하나의 인공지능이 원하는 답을 주지 못하면 다른 인공지능으로 옮겨가 시도해보는 경험을 하면서, 자연스럽게 여러 도구를 비교하고 각각의 장단점을 스스로 정리해보는 기회로 이어졌다.

마지막으로, 팀 단위로 계정을 돌려 쓰는 과정에서 사용량 제한이라는 다소 실질적이고 현실적인 제약도 함께 경험했다. 무료 계정에서는 질문 횟수나 기능에 한도가 있다는 것을 직접 확인하면서, 언제 질문을 아껴 써야 하는지, 어느 순간 다른 계정이나 다른 서비스로 넘어가야 하는지를 팀원들과 상의해가며 판단해야 했다. 이는 단순히 인공지능을 쓰는 법을 넘어, 제한된 자원을 팀 안에서 어떻게 효율적으로 나눠 쓸지 고민해보는 경험이 되었다.

11. 실제 테이터

1. 케플러의 법칙

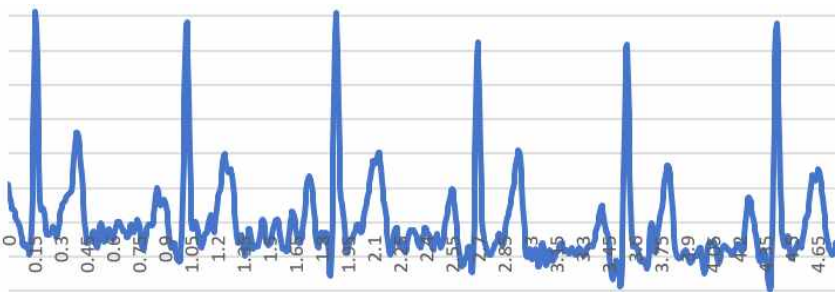
1.0	0.0553	0.3870	0.0880	1.0	0.000174	0.3870	0.0880	0.54	0.50	107.270	1089.0
1.0	0.815	0.7233	0.2247	1.0	0.00256	0.7233	0.2247	1.33	1.41	38.235	143.1
1.0	1.0	1.0	0.3652	1.0	0.00315	1.0	0.3652	0.88	0.82	113.769	930.0
1.0	0.107	1.5234	0.6870	1.0	0.000338	1.5234	0.6870	3.06	1.97	131.352	675.5
1.0	317.83	5.2045	4.331	1.0	1.0	5.2045	4.331				
1.0	95.16	9.5822	10.747	1.0	0.299	9.5822	10.747				
1.0	14.54	19.2012	30.589	1.0	0.0457	19.2012	30.589				
1.0	17.15	30.0475	59.800	1.0	0.0540	30.0475	59.800				
				0.33	0.018	0.0505	0.0072004				
				0.33	0.012	0.125	0.02814				
				0.33	0.008	0.213	0.06224				
				0.33	0.008	0.156	0.039026				
				0.33	0.014	0.549	0.2562				
				0.08	0.0027	0.0111	0.0015109				
				0.08	0.0043	0.0152	0.0024218				
				0.08	0.0013	0.0214	0.0040496				
				0.08	0.002	0.0282	0.0060996				
				0.08	0.0021	0.0371	0.0092067				
				0.08	0.0042	0.0451	0.0123529				
				0.08	0.086	0.063	0.018767				

2. 쌍성의 절대등급, 질량과 광도의 관계

1-A	23.8	-4.2	1.37657696	3.61
2-A	16.7	-3.45	1.22271647	3.31
3-A	15.6	-3.9	1.19312460	3.49
4-A	11.8	-3.15	1.07188201	3.19
5-A	9.2	-2.1	0.96378783	2.77
6-A	6.13	-1.7	0.78746047	2.61
6-B	5.25	-1.25	0.72015930	2.43
7-A	6.1	-1.55	0.78532984	2.55
7-B	6	-1.55	0.77815125	2.55
8-A	5.16	-1	0.71264970	2.33
8-B	4.6	-0.7	0.66275783	2.21
9-B	4.65	-0.3	0.66745295	2.05
10-B	4.12	-0.35	0.61489722	2.07
11-A	3.92	-0.37	0.59328607	2.08
12-A	3.61	-0.98	0.55750720	2.32
13-B	2.29	1.19	0.35983548	1.46
14-A	2.08	1.48	0.31806333	1.34
15-A	1.72	2.29	0.23552845	1.02

3. 심장박동 측정 엑셀 데이터의 일부:그래프는 제공하지 않고 숫자만 사용-이해를 돕기 위한 그래프

0	0.055678
0.005	0.038681
0.01	0.033705
0.015	0.023488
0.02	0.02017
0.025	0.019665
0.03	0.017526
0.035	0.014112
0.04	0.007404
0.045	0.003558
0.05	0.000385
0.055	-0.00058
0.06	-0.00325
0.065	-0.00719
0.07	-0.01618
0.075	-0.02277
0.08	-0.03068
0.085	-0.03363
0.09	-0.03459
0.095	-0.03378
0.1	-0.03253
0.105	-0.03318
0.11	-0.03806
0.115	-0.0433
0.12	-0.04808
0.125	-0.04262
0.13	-0.00875
0.135	0.035388
0.14	0.129026
0.145	0.19807
0.15	0.281539
0.155	0.307142
0.16	0.287068
0.165	0.239973
0.17	0.146479
0.175	0.088613
0.18	0.032142
0.185	0.017429
0.19	0.017598
0.195	0.02154



Ⅲ

정리

항목	1부 · 인공지능 스무고개
소재	감춰진 과학 데이터의 정체
핵심 리터러시	인공지능에게 좋은 질문을 만들어 가는 능력
난이도 유지 방법	인공지능이 풀기 어려운 문제 자체를 설계
팀·개인 구조	팀 전원 공동 수행
소요 시간	1차시(60분) 집중형
채점 방식	완료 여부 기반(만점 / 1점 감점)
시간 경과에 따른 취약점	1년 만에 인공지능 성능 향상으로 문제가 무력화됨

결론

인공지능 스무고개는 인공지능이 정답을 내어 놓지 못하는 상황을 계속 직면하게 하여 인공지능이 내놓은 답을 판단할 수 있는 판단력의 중요성을 인식시키고 인공지능에 대한 맹목적 의존성을 낮추는 것을 목적으로 하여 성공적인 교육적 효과를 달성하였다.

VI

운영상 유의점과 한계

1. 운영 시 유의점

- 팀별 인공지능 사용량(토큰) 제한을 고려해, 수업 전 팀원별로 계정을 여러 개 준비하도록 미리 안내해야 50분 안에 완료할 수 있다.
- 힌트를 줄 때 교사는 정답을 직접 알려주지 않고 "예/아니오"로만 응답해, 학생이 스스로 질문을 재구성하도록 해야 한다.
- 인공지능의 성능은 계속 향상되므로, 같은 자료를 다음 학년도에 재사용하기 전에는 교사가 먼저 직접 풀어 보고 난이도를 재검증해야 한다(Ⅲ장 참조).

위 인공지능을 활용한 빅데이터 활용 수업의 수행평가는 실제로 설계되고 숭의여자고등학교에서 진행된 것으로 다음 부록에는 이 자료를 참고하여 실시하려는 선생님들께 참고할 예시를 제공합니다.

부록 A

제1부 진행 대본

스무고개 힌트 시퀀스 예시

아래는 실제 수업에서 오간 힌트 흐름을 재구성한 예시다. 교사는 학생의 질문에 "예/아니오"로만 답하며 범위를 좁혀 준다.

학생 질문 이 데이터는 공학 분야의 데이터인가요?

교사 답변 아니요. 과학 데이터입니다.

학생 질문 화학 데이터인가요?

교사 답변 아니요. 물리·화학·생물·지구과학 중 하나입니다.

학생 질문 유체의 이동에 관한 데이터인가요?

교사 답변 아니요.

학생 질문 생물 데이터인가요?

교사 답변 세 자료 중 하나는 그렇습니다. 어느 표를 말하는지 정해서 다시 물어보세요.

학생 질문 (표 ③을 가리키며) 이건 사람 몸에서 나온 신호인가요?

교사 답변 예. 실험실에서 직접 측정한 신호입니다.

학생 질문 (표 ①을 가리키며) 이건 천체와 관련된 데이터인가요?

교사 답변 예.

학생 질문 (표 ②를 가리키며) 이것도 천체 관련인가요?

교사 답변 예. 다만 ①과는 측정하는 대상과 물리량이 다릅니다.

이런 방식으로 세 자료 각각의 큰 범주를 확정한 뒤에는, 표별로 "행성의 개수인가요?", "시간의 단위인가요?", "맥박인가요?"처럼 더 구체적인 질문과 답을 주고받으며 60분 안에 세 자료 모두의 정체(행성의 공전주기와 궤도 장반경, 쌍성의 밝기 변화, 심장박동 신호)와 그래프를 완성하도록 유도한다. 어떤 인공지능이 그래프를 그리는가 스스로 알아가게 유도한다.