AI 활용 사례 보고서

범위와 난이도만 정하면 2초: 정답을 계산으로 검산한 수학 테스트지를 한글·워드·PDF로

분야: 교육·학습을 위한 AI 활용 (교사 수업 준비)

테스트지는 매일 필요한데, 한 문제만 틀려도 안 됩니다

학원이나 학습지 선생님들은 일일 테스트, 주간 테스트, 월간 테스트를 계속 만듭니다. 배운 범위 안에서 내야 하고, 쉬운 문제와 어려운 문제를 섞어야 하고, 출판사 문제집에 있는 문제를 그대로 낼 수도 없습니다. 그래서 이 일만 맡는 조교나 문제 제작자를 따로 두는 곳도 있습니다.

이 분야가 유난히 까다로운 이유는 오타 때문입니다. 숫자 하나가 틀리면 답이 없는 문제가 되고, 그런 문제가 한 번 나가면 학생과 학부모는 테스트지 전체를 의심합니다. 문제를 사진으로 읽어 들이는 방식(OCR)이 오래 시도됐지만 이 '숫자 하나' 때문에 번번이 사람 손이 다시 들어갔습니다.

요즘 생성형 AI는 수학 문제를 꽤 잘 만듭니다. 그런데 문제와 답을 한꺼번에 맡기면 계산 실수가 섞여 나올 수 있고, 그럴듯하게 보여서 오히려 찾아내기 어렵습니다. 게다가 AI가 준 글을 한글 문서에 붙이면 수식이 수식으로 들어가지 않아 결국 수식 편집기로 다시 쳐야 합니다. 이대로 쓰면 빨라지는 게 아니라 검토할 일이 하나 더 생깁니다.

그래서 이렇게 만들었습니다

원칙은 하나로 잡았습니다. 문제는 AI와 만들되, 정답은 AI에게 묻지 않는다.

먼저 Claude와 함께 출제 틀을 만들었습니다. 고등 수학 Ⅰ, 수학Ⅱ, 미적분, 확률과 통계에서 37가지 유형을 골라 각각 쉬움·보통·어려움 세 단계로 나눴습니다. 테스트지를 뽑을 때는 프로그램이 틀 안에서 숫자를 새로 뽑기 때문에 같은 범위로 다시 뽑아도 다른 문제가 나옵니다. 가지고 있는 문제와 비슷한 문제나 더 어려운 문제가 필요할 때는 AI가 원본을 읽고 같은 유형으로 새로 씁니다.

정답은 검산 프로그램(sympy)이 따로 구합니다. 출제 과정에서 나온 답을 가져다 쓰지 않고, 문제지에 실제로 찍힌 식을 처음부터 다시 풉니다. 방정식이면 실근을 전부 찾고, 최대·최소 문제면 구간을 촘촘히 훑습니다. 이렇게 구한 값이 선택지 다섯 개 중 정확히 하나와 맞아야 통과입니다. 맞는 것이 없거나 둘이면 그 문제는 버립니다. '보기에서 옳은 것만 고르기' 문제는 ㄱ, ㄴ, ㄷ 문장을 하나씩 따로 계산해서 참·거짓을 가립니다.

문서는 한글(HWP), 워드(DOCX), PDF 가운데 골라 받을 수 있게 했습니다. 한글과 워드에서는 수식이 그림이 아니라 수식 개체로 들어가서, 선생님이 숫자 하나 바꾸고 싶을 때 바로 고칠 수 있습니다. 2단으로 할지, 가운데 선을 넣을지, 문제 사이를 몇 줄 띄울지, 선택지를 한 줄에 몇 개 놓을지도 정하는 대로 맞춥니다. 한 문제가 단이나 쪽 경계에서 반으로 갈라지지 않게 묶는 것도 넣었습니다. 테스트지를 받아 본 사람은 이게 얼마나 거슬리는지 압니다.

그림도 같은 원칙으로 넣습니다. 넓이 문제는 본문에 쓰인 식을 읽어 그래프를 그려 붙이는데, AI 그림 생성기는 쓰지 않습니다. 그런 도구는 눈금이나 길이 숫자를 계산하지 않고 그럴듯하게 그리기 때문입니다. 대신 식을 풀어 구한 교점으로 그리고, 색칠한 부분의 넓이를 따로 적분해서 그 문항의 정답과 같을 때만 붙입니다. 다르면 그림을 붙이지 않고 리포트에 남깁니다.

마지막으로 만든 파일을 다시 읽어서 처음 정한 문제 내용과 글자, 수식을 대조합니다. 하나라도 빠지거나 바뀌었으면 내보내지 않습니다. 한글 파일은 한글 프로그램에서 직접 열어 쪽마다 눈으로도

봤습니다.

실제로 돌려 본 결과

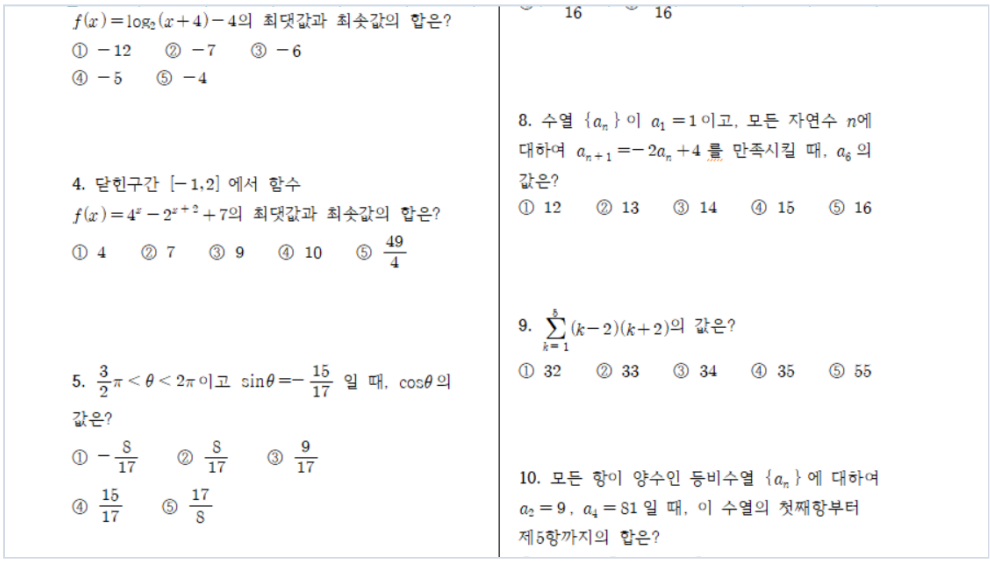
'수학Ⅱ 미분, 10문항, 쉬움 3개·보통 4개·어려움 3개'로 일일 테스트를 요청했더니 문제지, 정답·해설지, 검산 리포트가 2초 만에 나왔습니다. 같은 조건으로 한 번 더 뽑았을 때 10문항 모두 앞의 것과 다른 문제였습니다.

'수열과 적분, 20문항, 쉬움 6·보통 8·어려움 6' 주간 테스트를 한글, 워드, PDF 세 형식으로 한꺼번에 요청했을 때는 6초가 걸렸습니다. '미분과 적분, 20문항' 월간 테스트를 그림까지 넣어 세 형식으로 요청했을 때는 7초가 걸렸고, 넓이 문제 3개에 그래프가 자동으로 붙었습니다. 함께 첨부한 견본이 이때 나온 파일입니다.

고3 수학 40문항을 만들었을 때는 40문항 전부 검산을 통과했고, 파일 대조에서 빠지거나 바뀐 글자가 없었습니다.

검산 리포트에는 문항마다 정답, 다시 계산한 값, 어떤 방법으로 검산했는지가 적혀 있습니다. 선생님이 의심 가는 문제가 있으면 그 줄만 보면 됩니다.

만들면서 고친 것도 있습니다. 처음에는 계산 결과의 근삿값을 분수로 바꿔 주는 기능을 썼는데 값이 정확하지 않은 경우가 있어 빼고 정확한 계산으로 바꿨습니다. 해설지의 긴 수식이 단 너비를 넘어 잘리는 것은 화면을 직접 보다가 발견해서, 긴 수식을 등호에서 나눠 쓰도록 고쳤습니다. 프로그램이 제대로 도는지 보는 자동 점검을 233개 만들어 두고 고칠 때마다 다시 돌립니다.



$f(x)=\log_2(x+4)-4$의 최댓값과 최솟값의 합은?
① -12 ② -7 ③ -6
④ -5 ⑤ -4

4. 닫힌구간 $[-1,2]$에서 함수 $f(x)=4^x-2^{x+2}+7$의 최댓값과 최솟값의 합은?
① 4 ② 7 ③ 9 ④ 10 ⑤ $\frac{49}{4}$

5. $\frac{3}{2}\pi<\theta<2\pi$이고 $\sin\theta=-\frac{15}{17}$ 일 때, $\cos\theta$의 값은?
① $-\frac{8}{17}$ ② $\frac{8}{17}$ ③ $\frac{9}{17}$
④ $\frac{15}{17}$ ⑤ $\frac{17}{8}$

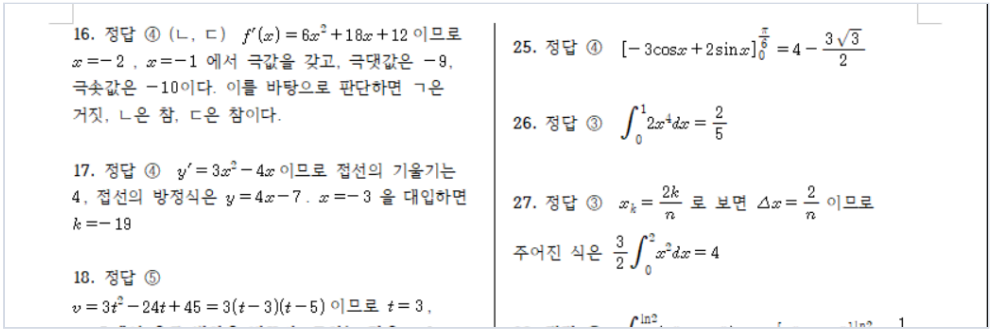
16 16

8. 수열 $\{a_n\}$이 $a_1=1$이고, 모든 자연수 n에 대하여 $a_{n+1}=-2a_n+4$를 만족시킬 때, a_6의 값은?
① 12 ② 13 ③ 14 ④ 15 ⑤ 16

9. $\sum_{k=1}^{8}(k-2)(k+2)$의 값은?
① 32 ② 33 ③ 34 ④ 35 ⑤ 55

10. 모든 항이 양수인 등비수열 $\{a_n\}$에 대하여 $a_2=9$, $a_4=81$일 때, 이 수열의 첫째항부터 제5항까지의 합은?

그림 1. 만들어진 문제지를 한글에서 연 화면. 수식은 한글 수식 개체입니다.



16. 정답 ④ (ㄴ, ㄷ) $f'(x)=6x^2+18x+12$이므로 $x=-2$, $x=-1$에서 극값을 갖고, 극댓값은 -9, 극솟값은 -10이다. 이를 바탕으로 판단하면 ㄱ은 거짓, ㄴ은 참, ㄷ은 참이다.

17. 정답 ④ $y'=3x^2-4x$이므로 접선의 기울기는 4, 접선의 방정식은 $y=4x-7$. $x=-3$을 대입하면 $k=-19$

18. 정답 ⑤
$v=3t^2-24t+45=3(t-3)(t-5)$이므로 $t=3$,

25. 정답 ④ $[-3\cos x+2\sin x]_0^{\frac{\pi}{6}}=4-\frac{3\sqrt{3}}{2}$

26. 정답 ③ $\int_0^1 2x^4dx=\frac{2}{5}$

27. 정답 ③ $x_k=\frac{2k}{n}$로 보면 $\Delta x=\frac{2}{n}$이므로
주어진 식은 $\frac{3}{2}\int_0^2 x^2dx=4$

그림 2. 함께 나오는 정답 및 해설지.

아직 안 되는 것

솔직하게 적습니다. 그림은 이제 막 시작한 단계입니다. 자동으로 붙는 것은 넓이 문제의 그래프뿐이고, 삼각형은 견본으로 확인한 정도입니다. 그 밖의 그림은 선생님이 그림 파일을 주면 넣는 방식입니다. 기하, 단답형, 서술형, 증명 문제는 아직입니다.

화면으로 확인한 것은 한글 2010과 PDF입니다. 워드 파일은 수식 개수와 글자가 맞는지 프로그램으로만 확인했습니다.

그리고 검산이 보장하는 것은 '정답이 맞다'까지입니다. 그 문제가 우리 반 학생에게 좋은 문제인지는 여전히 선생님이 판단할 몫입니다. 아직 학원 현장에서 선생님이 써 본 기록은 없고, 위 결과는 제가 직접 만들어 확인한 것입니다.

기대하는 효과

테스트지를 만들려고 조교나 문제 제작자를 따로 두던 비용과 시간을 줄일 수 있다고 봅니다. 선생님은 문제를 옮겨 치고 답을 맞춰 보던 시간을 수업 준비와 학생 지도에 쓸 수 있습니다.

사람이 문제를 내면 손에 익은 숫자와 익숙한 꼴로 기울기 마련입니다. 프로그램은 조건 안에서 숫자를 고르게 뽑으니, 출제자의 주관이나 경험이 끼어들지 않은, 평소 잘 내지 않던 꼴의 문제가 나올 수 있습니다.

같은 범위에서 난이도 비율만 바꿔 여러 벌을 바로 뽑을 수 있으니 반별, 수준별 테스트를 따로 내기도 쉬워집니다.

다만 이 세 가지는 아직 현장에서 재 본 숫자가 아니라 제가 기대하는 효과입니다.

다른 분들이 따라 하려면

수학 문제지가 아니어도 순서는 그대로 쓸 수 있습니다. AI가 만들고, AI가 아닌 다른 방법으로 검사하고, 통과한 것만 쓴다. 견적서나 정산표처럼 숫자가 틀리면 안 되는 문서에 AI를 쓸 때, 엑셀 수식이든 원본 대조든 한 번 더 확인하는 단계를 두면 됩니다.

제가 쓴 계산·문서 도구(sympy, python-hwpx, pandoc)는 모두 무료로 공개된 것입니다.

책임 있는 사용을 위해 지킨 것

문제는 프로그램이 숫자를 새로 뽑아 만들고, 시중 문제집의 문제를 옮겨 오지 않습니다. 비슷한 문제를 만들기 위해 기준 문제를 받을 때는 사용 권한이 있는 자료만 씁니다. 학생 이름이나 성적 같은 개인정보는 쓰지 않습니다.

사용한 AI는 Claude(Claude Code)이고, 프로그램 제작과 이 보고서·영상의 정리에 함께 사용했습니다. 영상의 음성은 AI 음성(ElevenLabs)입니다.