AI는 15문제 중 1개만 맞힌 카드 할인 계산, 계산은 코드에 맡기는 챗봇을 직접 만들어 15개 모두 맞혔어요

공모 분야: 생활 속 AI 활용 · 활용한 AI 도구: Claude Code, Claude, ChatGPT(OpenAI GPT-4o-mini)

① 어떤 상황에서 AI를 활용했나요?

제 지갑에는 카드가 9장(체크카드 3장, 신용카드 6장) 있습니다. 카드마다 할인되는 곳, 전월에 써야 하는 금액, 한 달 할인 한도가 다 다릅니다. 예를 들어 한 카드는 스타벅스 50% 할인이지만 하루 1번, 한 달 2번까지만 되고, 영화관은 또 다른 카드가 35%입니다. 계산대 앞에서 이걸 다 기억할 수 없어서, "어디서 얼마 쓰면 어떤 카드가 이득인지" 알려 주는 챗봇을 만들기로 했습니다.

처음에는 AI에게 그냥 물어보면 될 줄 알았습니다. 그런데 "오후 7시 CGV에서 영화 18000원"을 물었더니 AI(GPT)가 카드 세 장 모두 "18,000원 할인"이라고 답했습니다. 영화가 공짜라는 뜻이죠. 상품설명서로 계산한 정답은 6,300원(18,000원 × 35%)이었습니다.

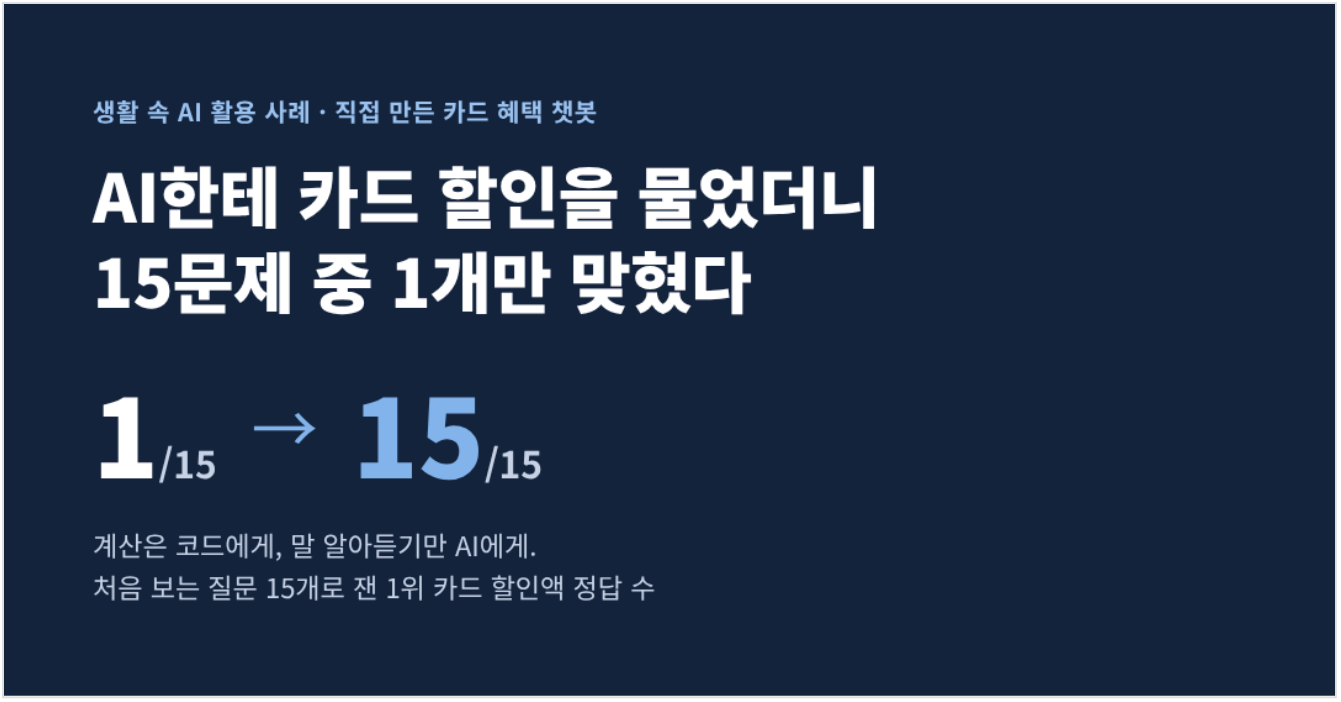


사진 1. 처음 보는 질문 15개로 잰 1위 카드 할인액 정답 수



사진 2. AI의 오답(왼쪽)과 상품설명서로 계산한 정답(오른쪽)

② 어떤 AI를 어떻게 활용했나요?

- Claude Code: 챗봇 코드를 작성·수정·테스트했습니다. 저는 무엇을 만들지, 어디가 틀렸는지, 어떻게 바꿀지를 정했습니다. (9월 29일 시작, 10월 12일 최종 점검까지 약 2주)
- ChatGPT(GPT-4o-mini): 챗봇 안에서 코드가 모르는 가게 이름이 나오면 "이 가게일 것 같다"는 후보만 냅니다. AI 단독 답변과 비교 측정에도 썼습니다.
- Claude: 시험 문제의 정답을 먼저 계산했고, 그 정답을 제가 상품설명서 원문과 직접 대조했습니다.

먼저 카드사 상품설명서 PDF 12개를 읽어 카드 9장, 혜택 65개를 같은 틀의 데이터로 정리하고, 혜택마다 "설명서 몇 쪽"이라는 근거를 달았습니다.

그리고 AI에게 맡기는 일의 범위를 바꿨습니다.

- 처음: AI가 질문 이해, 카드 조건 판단, 금액 계산, 답변까지 전부 했습니다. 같은 질문인데 실행할 때마다 해석이 달라지기도 했습니다.
- 최종: 코드가 질문을 정해 둔 칸(가게·금액·수량·시간·요일·결제수단)에 채우고, 전월 실적·한도·남은 횟수를 따져 금액을 계산합니다. AI는 코드가 모르는 가게 이름의 후보만 내고, 그 이름이 질문에 글자 그대로 있을 때만 씁니다.

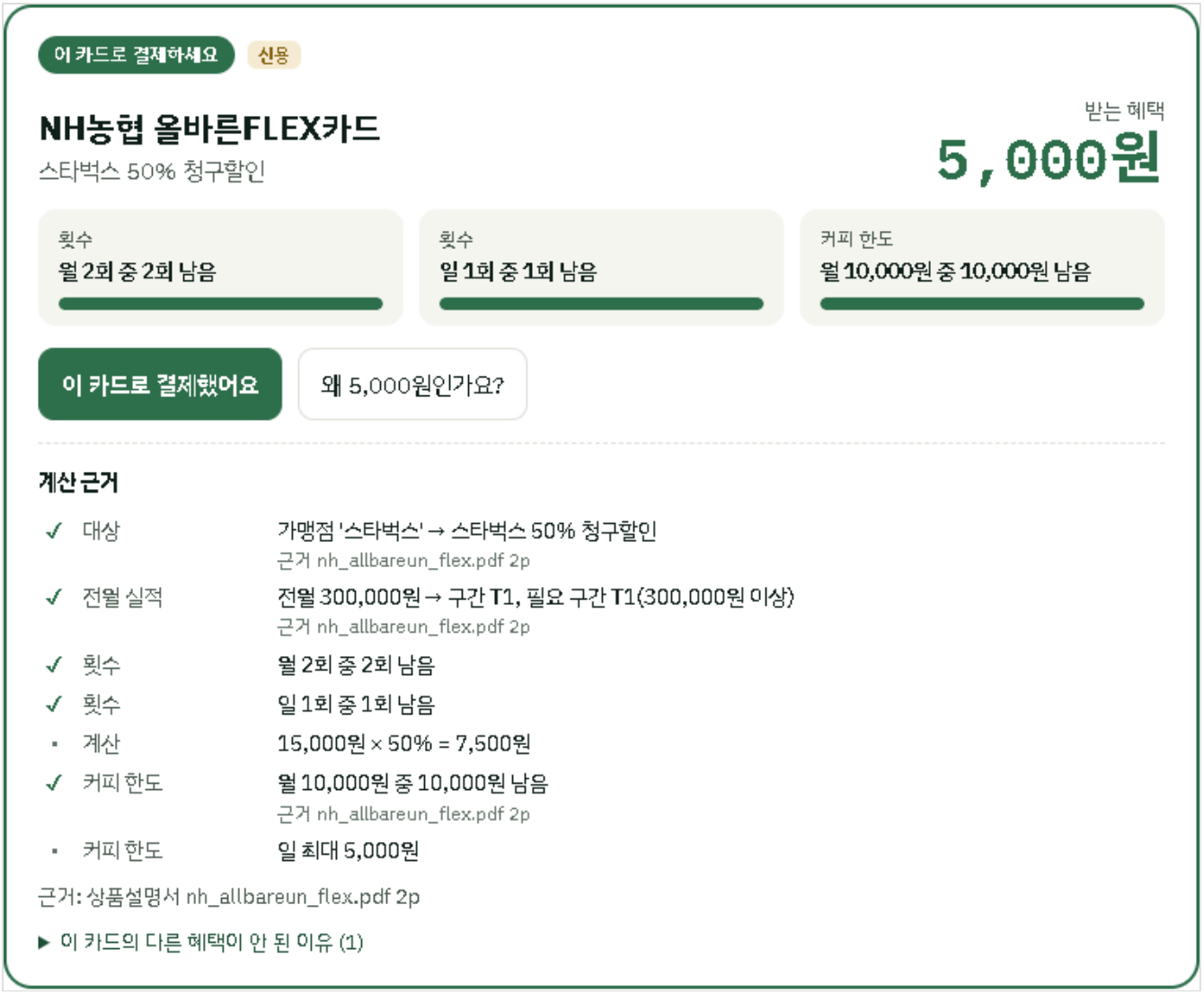


사진 3. 직접 만든 챗봇의 답변 화면: 추천 카드, 남은 횟수·한도, 설명서 쪽수가 달린 계산 근거



사진 4. AI에게 맡기는 일의 범위를 바꾼 구조 전/후

③ 활용 결과 어떤 변화가 있었나요?

개발 중 한 번도 쓰지 않은 가게(CGV, 에버랜드, 올리브영, 요기요, 고속버스 등)로 시험 문제 15개를 만들고, 문제와 정답을 먼저 저장해 잠근 뒤 한 번만 측정했습니다. 1위 카드 할인액을 맞힌 개수입니다.

- AI에게 카드 자료 없이 그냥 물음: 15개 중 1개
- AI에게 카드 자료를 주고 계산까지 맡김: 15개 중 5개
- 계산은 코드, AI는 가게 이름 후보만 (최종 챗봇): 15개 중 15개

개발하면서 쓴 100문항에서도 AI 단독 11개 → AI 계산 42개 → 최종 챗봇 100개로 바뀌었습니다. 측정할 때마다 새로 만든 문제에서도 2차 15개 중 15개, 3차 40개 중 39개를 맞혔습니다. 틀린 1개는 "아웃백"의 '백'을 숫자 100으로 읽은 경우인데, 점수를 올리려고 고쳐서 다시 재지 않고 실패 사례로 남겼습니다. 정답 70개는 10월 12일 제가 설명서 원문과 직접 대조했고, 고칠 것은 없었습니다.

이제 결제 전에 "오후 2시 스벅 15000원"이라고만 치면, 어떤 카드로 내야 하는지, 이번 달에 몇 번·얼마 더 받을 수 있는지가 "15,000원 × 50% = 7,500원 → 커피 하루 한도 5,000원"처럼 계산 근거와 설명서 쪽수와 함께 나옵니다.

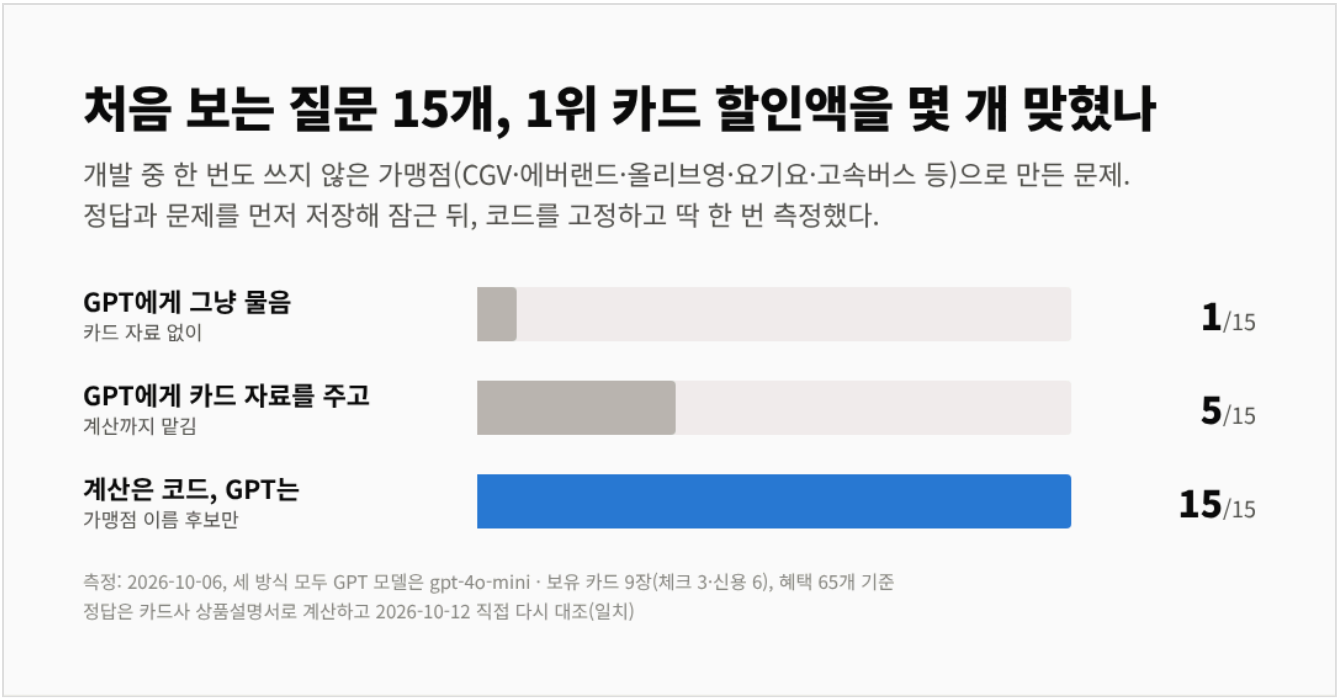


사진 5. 세 가지 방식의 정확도 비교 (처음 보는 15문항)

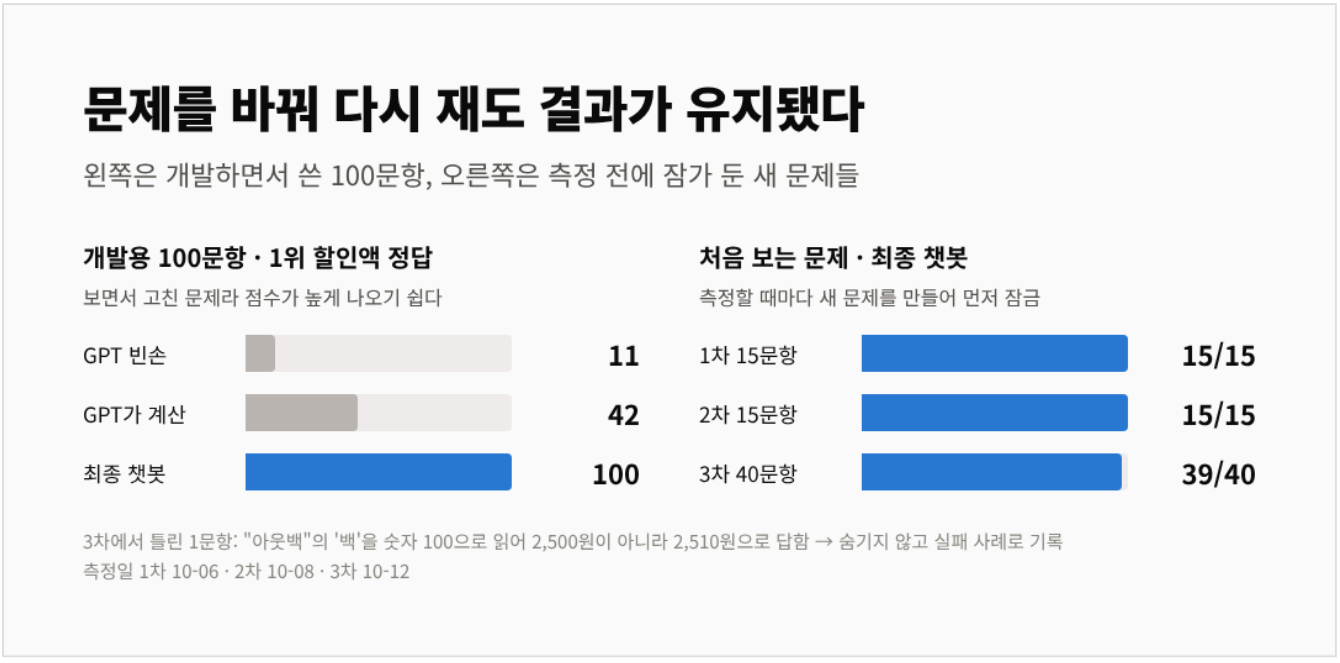


사진 6. 개발용 100문항과 새로 잠근 문제들로 다시 잰 결과

④ 나만의 방식 또는 개선 포인트는 무엇인가요?

- 계산은 코드에, 말 알아듣기만 AI에게: 틀리면 손해인 숫자는 AI에게 맡기지 않았습니다.
- 모든 답에 근거: 혜택마다 설명서 쪽수를 달아, 답이 맞는지 바로 확인할 수 있게 했습니다.
- 일부러 헷갈리게 물어보기: 헷갈리는 질문 62개로 시험했더니 오류가 33건 나왔습니다. "이마트24 12000원"을 241만 2천 원으로, "아마존 직구 5만원"을 95만 원으로('구'를 숫자 9로) 계산하는 식이었습니다. 모두 고친 뒤 0건이 됐습니다.

- 처음 보는 문제로 다시 재기: 고친 문제만 다시 돌리면 점수는 늘 좋게 나옵니다. 새 표현으로 다시 시험하자 30개 중 26개만 맞아서, 측정할 때마다 새 문제를 만들어 미리 잠가 두었습니다.
- 모르면 적게, 모르면 묻기: 혜택이 겹칠 때 설명서에 "같이 받을 수 있다"고 쓰여 있지 않으면 큰 혜택 하나만 계산합니다. "일본에서 1만엔"처럼 원화 금액을 모르면 짐작하지 않고 되묻습니다.



일부러 헷갈리게 물어서 찾아낸 오류들

1차 공격 테스트 62문항, 두 가지 해석 방식에서 오류 33건(24건 + 9건) → 모두 고친 뒤 0건

질문	챗봇의 잘못된 계산	원인
이마트24 12000원	241만 2천원	띄어쓰기를 지운 뒤 숫자를 찾아 "24"와 금액이 붙음
아마존 직구 5만원	95만원	"직구"의 '구'를 숫자 9로 읽음
1.5만원	5만원	소수점을 처리하지 못함
KB Pay **말고** 실물카드로	KB Pay로 계산	"말고" 같은 부정 표현을 놓침
스타벅스 카드 **충전**	할인된다고 답함	선불 충전은 할인 제외라는 조건을 놓침
케이패스로 스벅	KB Pay로 짐작	말하지 않은 결제수단을 AI가 채워 넣음

출처: 개발 기록(2026-10-01). 고친 문제만 다시 돌리면 늘 통과해서, 그 뒤로는 처음 보는 표현으로 다시 공격했다(3차 첫 실행 26/30).

사진 7. 일부러 헷갈리게 물어서 찾아낸 오류들

⑤ 다른 사람도 따라 할 수 있나요?

네. 챗봇을 만들지 않아도 AI에게 숫자가 걸린 질문을 할 때 바로 쓸 수 있습니다.

1. 할인액·이자·세금처럼 틀리면 손해인 숫자는 엑셀이나 계산기로 다시 셈합니다. AI에게는 "어떤 조건을 따져야 하는지"를 묻는 게 더 안전합니다.

2. "설명서 몇 쪽에 있는 내용인지 같이 적어 줘"라고 근거를 함께 요구합니다.

3. 내가 답을 아는 질문 몇 개로 먼저 시험해 보고, 그 AI를 어디까지 믿을지 정합니다.

4. "말고", "1.5만 원", "이마트24"처럼 실제로 쓰는 말투로 헷갈리게 물어봐야 진짜 약점이 보입니다.

챗봇을 직접 만들고 싶다면, 코드를 몰라도 Claude Code 같은 AI 코딩 도구에 "무엇을 만들지"와 "정답 예시"를 주고 시작할 수 있습니다. 중요한 건 AI에게 무엇을 맡기고 무엇을 사람이 확인할지 정하는 것이었습니다.

※ 카드 정보는 제가 보유한 카드의 상품설명서 기준이며, 특정 카드사·상품을 홍보하지 않습니다.

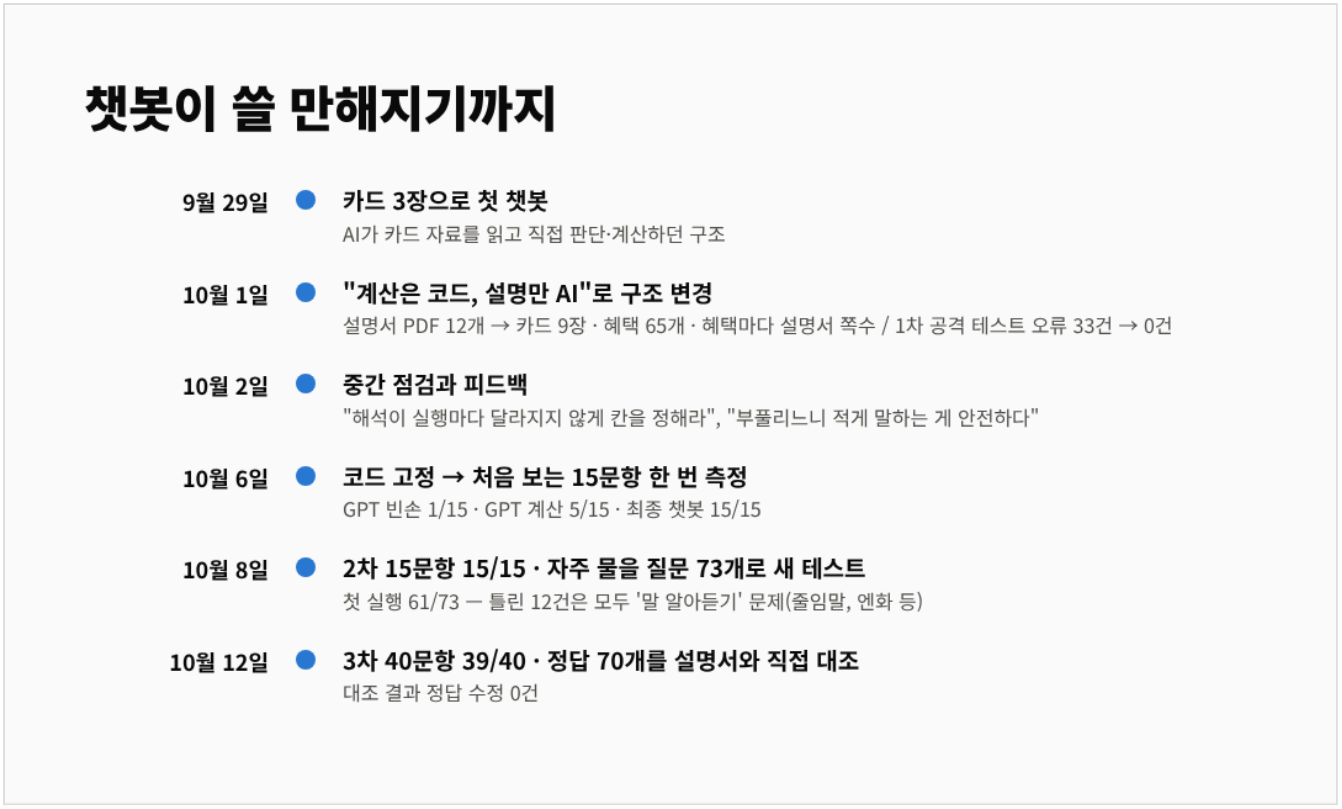


사진 8. 챗봇이 쓸 만해지기까지 (9월 29일~10월 12일)