

FIELD NOTE 09

The FDA Narrowed Keytruda's Label in May 2025. Seventeen Months Later, All Four AI Models Still Gave the Old One.

Asked on 22 October 2026 about a patient whose tumour scores below 1 on PD-L1, each of four AI models stated in two or three of six answers that first-line Keytruda with chemotherapy is FDA-approved regardless of PD-L1, as it was before the FDA restricted the indication. In the same pre-specified studies, two models told US families that people with two copies of APOE ε4 are excluded from Leqembi's US approval, which they are not, and Claude told patients in the EU that Leqembi can be started without the genetic test the EU requires. One day of answers establishes that the errors were made, not how often; a confirmatory study is testing all three, the Keytruda one in oesophageal cancer only.

Ozue C. Ezeife

29 October 2026

Oncology, Alzheimer's disease and type 1 diabetes · United States, European Union ·
4 engines · 21 min read

Until 22 May 2025, Keytruda with chemotherapy was approved in the United States as a first-line treatment for advanced oesophageal cancer whatever the tumour's PD-L1 level, and for HER2-negative gastric cancer on the same terms. That day the FDA restricted both indications to tumours with a PD-L1 combined positive score (CPS) of 1 or more [1]. Seventeen months later, on 22 October 2026, AlphaCitation asked four AI models whether a patient below that line is FDA-approved for the combination. Each of the four stated the

indication as it stood before that date, under which the patient is covered, in two or three of its six answers: 10 of 24 in all. Eight of the ten answered yes; two of Claude's three answered no, then said the approval applies regardless of PD-L1 expression.

Claude Haiku 4.5 put it plainly. Keytruda with chemotherapy, it said, "is FDA-approved for first-line treatment of esophageal and gastroesophageal junction carcinoma, regardless of histology or PD-L1 expression", and so "yes" for a tumour with a CPS below 1. GPT-4o mini said the FDA's approval for HER2-negative gastric cancer "does not specify a requirement for PD-L1 expression levels". Both sentences describe the label as it stood before May 2025. The current label says otherwise [2].

The Keytruda result is one of four pre-specified studies that asked the same question of the same four models, GPT-4o mini, Claude Haiku 4.5, Gemini 2.5 Flash and Sonar, with web search on: does the answer give the label that applies, in the market asked about, on the day asked? For each medicine the US and EU labels differ, and the questions were written so that the right answer depends on which one applies.

Lecanemab (Leqembi) is the sharpest case. Its US label lets people with two copies of the APOE ε4 allele be treated; its EU authorisation excludes them [3] [4]. Asked about US patients, GPT-4o mini said such people are excluded in 5 of 18 answers and Claude in 3 of 18. Told that a mother in Ohio carries two copies, GPT-4o mini answered that "she would not be eligible for Leqembi treatment based on current criteria". The source it cited was the European Medicines Agency's page for the medicine. In the other direction, Claude told people in France and Belgium, in 4 of 6 answers (3 confirmed on adjudication), that treatment can begin without a genetic test, which is how the US label puts it; the EU authorisation makes the genotype a condition of treatment.

Teplizumab (Tzield) is the counter-example. In April 2026 the FDA lowered its age floor from 8 to 1 [5]. Six months later, not one of the 60 answers about US children under 8 gave the old floor as the current one.

Each of these errors is about eligibility: who may receive a medicine on its label, in the country where the question is asked. A family, a prescriber or a payer who acted on one would be applying another country's rules, or rules that have since changed.

These are counts from one day, not established rates. Each study was planned for four days (the companion study, two); collection stopped after the first day (for the companion study, after two of that day's three rounds), when the process that collected the answers ended, and each study is analysed as a one-day study. The plans fixed in advance when a count becomes a finding: an error rate is called established only when its 95% interval, widened because repeated answers to one question are not independent, has a lower bound of at least 10%. At these sizes no primary count clears that bar. A separate confirmatory study, frozen and hashed on 28 October 2026 after its smoke test and before its first collection call, is testing the three findings above from 29 to 31 October 2026 (the Keytruda one in

oesophageal cancer only) with new wordings and eight models: these four and one current-generation model from each of their developers, GPT-5.5, Claude Sonnet 5.5, Gemini 3.8 Flash and Sonar Pro. Its result will be published whatever it shows.

Keytruda: the label the models gave was the one before May 2025

The study asked twenty questions, three times each: sixteen about first-line Keytruda with chemotherapy in oesophageal squamous cell carcinoma and two about HER2-negative gastric cancer, in the United States and the European Union, and two about its registration in Saudi Arabia and the UAE. The two labels set different thresholds. The US label requires a CPS of 1 or more; the EU's requires 10 or more [\[6\]](#). A patient at CPS 1 to 9 is therefore within the US indication and outside the EU one, and a patient below 1 is outside both.

Exhibit 1. Did the models state the superseded all-comers indication for first-line Keytruda with chemotherapy? 24 answers to 2 questions, 6 per model, 22 October 2026.

MODEL	STATED THE PRE-MAY 2025 INDICATION	OESOPHAGEAL QUESTION	GASTRIC QUESTION
GPT-4o mini	2 of 6	0 of 3	2 of 3
Claude Haiku 4.5	3 of 6	3 of 3	0 of 3
Gemini 2.5 Flash	2 of 6	1 of 3	1 of 3
Sonar	3 of 6	0 of 3	3 of 3

One question asked about oesophageal squamous cell carcinoma below CPS 1, the other about HER2-negative gastric cancer at CPS 0; both indications have required CPS 1 or more since 22 May 2025. Each answer counted here stated the indication as it stood before that date (coder A; coder B flags 9 of the 10) and was coded false on the question: eight answered yes, and two of Claude's three answered no, then said the approval applies regardless of PD-L1 expression. The plan fixed these questions as a control for over-affirmation, not as a primary measure. Source: per-answer codes, reproduced by the study's tally script.

The old label did not only produce wrong answers. It also produced right ones for a reason that is no longer true. Asked about a patient at CPS 1 to 9, who is within the current US indication, the models were mostly right, but 34 of the 60 answers stated the pre-2025 indication for all comers, 31 of them in answers coded correct: Sonar in 12 of 15, Claude in 11 (3 of them false), Gemini in 7 and GPT-4o mini in 4. Under that label the patient is covered too, so an answer built on it comes out correct. The study's primary measure counts the conclusion, and cannot tell the two apart; the flag on each answer can.

Claude also made the opposite error. In 4 of its 15 answers about a US patient at CPS 1 to 9 it said the patient is outside the US indication, a stricter threshold than the label sets. In the EU it said, in 3 of 6 answers, that a patient at CPS 15 or 20, inside the EU threshold, is outside the EU authorisation. Asked by an oncologist in France about a patient at CPS 20, it answered "no, Keytruda plus chemotherapy is not an EU-authorised first-line option", then quoted the EU's CPS ≥ 10 condition and said the patient's score "exceeds this threshold".

Eight of Claude's ten false claims in this study contradict themselves within the same answer, and the adjudicating panel agreed on all eight. Because Claude failed this EU control, the plan does not read its correct EU answers as knowledge of the EU label.

GPT-4o mini and Gemini each placed a CPS 1 to 9 patient within the EU indication once in 15 answers; GPT-4o mini said twice in 6 that a CPS 15 or 20 patient is outside it. Apart from the questions below CPS 1, Sonar made no false claim on either label; it still stated the pre-2025 indication in 12 of its 15 answers at CPS 1 to 9, and gave the US threshold as CPS 10 in 2 of its 6 answers at CPS 15 or 20.

Leqembi: one gene, two labels

The US and EU labels for lecanemab differ on genotype eligibility and pre-treatment testing. Both warn that people with two copies of APOE ϵ 4 have a higher rate of ARIA, the brain swelling and bleeding the treatment can cause. The US label lets them be treated, says testing should be done before the first dose and says that a patient who is not tested can still be treated [3]. The EU authorisation covers only non-carriers and people with one copy, says lecanemab is not indicated in people with two, and requires each patient's genotype to be registered before the first infusion [4].

Exhibit 2. Lecanemab and two copies of APOE ε4: answers that gave the other market's rule. 144 answers to 12 questions, 36 per model, 22 October 2026.

MODEL	US QUESTIONS: SAID PEOPLE WITH TWO COPIES ARE EXCLUDED	EU QUESTIONS: SAID PEOPLE WITH TWO COPIES ARE INCLUDED
GPT-4o mini	5 of 18	0 of 18
Claude Haiku 4.5	3 of 18	0 of 18
Gemini 2.5 Flash	0 of 18	0 of 18
Sonar	0 of 18	0 of 18

The study's primary measures, fixed before collection. Each cell counts answers that stated the false claim for that market, of 18. Source: per-answer codes, reproduced by the study's tally script.

The error ran in one direction only. No model carried the US scope into the EU. The US errors sat in two of the six US questions, and mostly in one: a family member asking whether a mother in Ohio with two copies is eligible. GPT-4o mini and Claude each said she is not in all three rounds. Claude's version was that she "does not meet the standard FDA approval criteria for Leqembi". GPT-4o mini also gave the exclusion in two of three answers to a direct question. To every other US question, both models answered correctly in every round.

The two models reached the same wrong answer by different routes. Each of GPT-4o mini's five false answers cited a page on the European Medicines Agency's website, and four of them describe the indication in the EU's own words, "non-carriers or heterozygotes" (two without the hyphen). None of Claude's three cited a regulator's page; two attributed the restriction, in words, to the FDA's approval criteria. The record shows what each answer cited, not which sentence came from which page.

A companion study asked the eligibility question in three voices, a patient's, a caregiver's and a physician's, and asked separately whether a genetic test is needed first. All four false US eligibility answers, two from GPT-4o mini and two from Claude, came to the caregiver's question. The test fixed in advance, the physician's question against the patient's, found

none in four answers to each, for every model, and so no evidence of a difference at this size.

Exhibit 3. Must an APOE ε4 test come before lecanemab? Answers that said it need not. 48 answers to 6 questions, 6 per model and market, 22 October 2026.

MODEL	US: SAID TESTING IS UNNECESSARY	EU: SAID TREATMENT CAN START WITHOUT TESTING
GPT-4o mini	0 of 6	0 of 6
Claude Haiku 4.5	0 of 6	4 of 6
Gemini 2.5 Flash	0 of 6	1 of 6
Sonar	0 of 6	0 of 6

US truth: testing is recommended, and an untested patient can still be treated. EU truth: the indication depends on genotype and the genotype is registered before the first infusion. These questions were reported descriptively, outside the companion study's primary measures. The counts are coder A's; the adjudicating panel confirmed 3 of Claude's 4 and did not decide the fourth. Source: per-answer codes and the adjudication record.

Claude's EU answers came to the patient's and the caregiver's questions, two of two each; the physician's question drew none. Asked by someone in France whether a test must come first, Claude headed its answer "Testing is Recommended but Not Mandatory", quoted the FDA label's sentence that untested patients "can still be treated", and said that "technically you could start Leqembi without testing", adding that "it's strongly advisable to get the APOE4 test first". All four answers also recommended the test; three were flagged as contradicting themselves, and the panel confirmed that contradiction in one and divided on two. The companion study collected two of the three rounds planned for its first day, 66.7% of that day and below the 90% completeness its plan requires, so no reading is declared on these counts. On the two rounds observed, Claude's count exceeded the numerical confidence-bound threshold. But the study did not meet its pre-specified completeness criterion, so no established-error finding is declared.

The companion study also asked about Legembi's registration in Saudi Arabia, in English and in Arabic. Claude, Gemini and Sonar each stated it in all 24 of their answers. GPT-4o mini stated it in all 12 English answers and in 9 of 12 Arabic ones; the other three made no statement or hedged. The pre-specified test found no significant difference ($p = 0.25$ before correction for the nine tests in its family). Asked about aducanumab in the EU and about brexpiprazole for agitation in Alzheimer's dementia in both markets, no model made a false claim in any of its 12 answers, nor about rivastigmine, the control.

Tziold: the change the models did pick up

On 20 April 2026 the FDA extended teplizumab's indication, delaying Stage 3 type 1 diabetes in people with Stage 2 disease, from age 8 down to age 1 [5]. In the EU, where it is authorised as Teizeild, the floor is still 8 [7]. Three of the study's five questions in each market asked about a child aged 3, 4 or 5; two asked for the youngest approved age or the ages covered.

Exhibit 4. Can a child under 8 with Stage 2 type 1 diabetes receive teplizumab? 120 answers to 10 questions, 30 per model, 22 October 2026.

MODEL	US QUESTIONS: PLACED THE CHILD WITHIN THE US INDICATION	EU QUESTIONS: SAID THE CHILD IS WITHIN THE EU INDICATION
GPT-4o mini	13 of 15	0 of 15
Claude Haiku 4.5	15 of 15	0 of 15
Gemini 2.5 Flash	15 of 15	0 of 15
Sonar	15 of 15	0 of 15

The study's primary measures, fixed before collection. Source: per-answer codes, reproduced by the study's tally script.

No answer to a US question gave 8 as the current US floor, and no model carried the US floor of 1 into the EU. GPT-4o mini's two false answers were both to the question about a 3-year-old, and both contradict themselves: each states the US indication from age 1, then says that a 3-year-old "falls between the approved age ranges". The old floor did surface in five of GPT-4o mini's answers to EU questions, where it was given as the US rule with no mention of the April 2026 change. And ten answers to EU questions, five from GPT-4o mini and five from Gemini, dated the European Commission's authorisation to 12 January 2026; the decision is dated 8 January.

Why this change reached the answers and Keytruda's did not, the study cannot say. Both are FDA actions, Tzield's six months old and Keytruda's seventeen. Tzield's widened the indication; Keytruda's narrowed it.

What the records establish, and what they cannot

Every count in this note is a count of answers from one day, and answers to the same question are not independent: in the lecanemab study, all of Claude's false answers and most of GPT-4o mini's were to one question. That is why the plans widen each interval for repetition before reading it, and why no primary count is called an established rate. GPT-4o mini's 5 of 18 on lecanemab would clear the 10% bar if its answers were independent; widened for repetition, its interval runs from 7.9% to 63.2%. Claude's 4 of 15 stricter-threshold answers on Keytruda behave the same way. The one count declared above the bar is Claude's 3 of 6 on the EU Keytruda control (an interval of 18.8% to 81.2%); Claude's 4 of 6 on EU testing also clears it (20.8% to 93.9%) but is not declared, because the companion study fell below its completeness criterion. The declared reading rests on the two control questions showing no repetition effect. Had they shown the effect seen in the neighbouring questions, it would not clear the bar.

At these sizes, the plans cannot establish reliability either. Even zero observed errors in 18 answers is consistent with a true error rate as high as 17.6%. An error not observed on this day may still occur.

What the records do establish is that each of these answers was given. On 22 October 2026, each of the four models, asked through its developer's API with web search on, stated a version of Keytruda's indication that the FDA had replaced seventeen months earlier; two of them told US families that a medicine's US label excludes patients it does not exclude; and one told EU patients that a test the EU makes a condition of treatment is optional. The cause is not established. The pattern is consistent with answers that mix two markets' labels, or a label's current and earlier versions, but the record shows what each answer said and cited, not why.

Under AlphaCitation's Evidence-Authority Standard, a claim about who may receive a medicine in a market is settled only by that market's authorisation record and authorised

label, in force on the claim's date [8]. Every answer here was coded that way, against the labels read on the day of collection. The confirmatory study now running asks the three questions again, the Keytruda one in oesophageal cancer only, with new wordings, on three further days, of eight models.

For a regulated medicine, the right answer in one country can be the wrong answer in another. An answer has to say which label it is applying, and apply the one in force.

How we know this

Question	Do four AI models state a medicine's indication by the label of the market asked about, where the US and EU labels differ, and as the label stands on the day asked?
Studies	S2, lecanemab and APOE ε4 homozygotes (US within, EU outside); S2C, its companion: the question's voice, testing before treatment, Saudi registration in English and Arabic, and three comparators; S3, teplizumab for children under 8 with Stage 2 type 1 diabetes (US within since 20 April 2026, EU outside); S4, first-line pembrolizumab with chemotherapy in oesophageal squamous cell carcinoma by PD-L1 band (US CPS ≥ 1 since 22 May 2025, EU CPS ≥ 10).
Plans	Each hashed before a smoke test, then frozen and hashed after it and before its first collection call on 22 October 2026, with the question wordings unchanged between the two hashes; plan sha256: S2 9d6c532c..., S3 3461e1b3..., S4 9f0d4be8..., S2C b2340457..., each with its sealed wordings' hash in the released freeze record. Changes after the freeze are numbered deviations in each study's internal result record; none changes a count in this note.
Models	gpt-4o-mini through OpenAI's API with web search; Claude Haiku 4.5 (claude-haiku-4-5) with web search; Gemini 2.5 Flash with Google Search grounding; Sonar (Perplexity). The instrument's fixed settings; no system instruction; each model except Sonar, which always searches, chose whether to search.
Questions	20 wordings each in S2, S3 and S4, and 32 in S2C, written in the voices of patients, families, physicians and payers, and directly; the US and EU wordings mirrored, plus registration wordings for Saudi Arabia and the UAE. Each is identified by its hash; the texts are withheld, as in Field Notes 07 and 08.
Schedule	22 October 2026. S2, S3 and S4: three rounds, 240 answers each. S2C: two of the three rounds planned for its first day, 256 answers. 976 answers in all, every call answered.

Deviation	The plan was four days (S2C: two); collection stopped after day 1 when the runner process ended; each study is analysed as a one-day study. In S2C, day 1 itself stopped after two of three rounds, so its completeness, 66.7% of the day, is below its 90% criterion and no MATERIAL or NOT MATERIAL reading is declared in it.
------------------	--

Coding	Every answer read in full and coded on its target, in the market named, as false, correct, hedged or not addressed, by rules fixed before collection, with flags for the earlier label, self-contradiction and other wrong regulatory facts. Two model coders coded every answer, each blind to the other and to the model, the round and the order of collection; they agreed on 968 of 976 codes (Cohen's kappa 0.92 to 0.99 by study).
---------------	---

Adjudication	Every false code, every coder disagreement on the measured questions and every self-contradiction flag, 49 answers, was adjudicated before publication by a panel of AI models from Anthropic, OpenAI and Google, each blind to the others, to the coders and to the model that wrote the answer. A code counted only if every judge on the panel for that answer gave it with a sentence quoted verbatim from the answer. 24 answers of known code, written from the labels, were mixed in unmarked; the panel decided all 24 and got all 24 right. It confirmed the first coder's code on 47 of the 49 and decided neither way on 2, one of them one of Claude's four EU testing answers. Of the 21 self-contradiction flags the first coder placed among them, it confirmed 15, divided on 3 and found no contradiction in 3. The six coder disagreements on the descriptive Gulf classes decide no reading and are reported as coded.
---------------------	---

Truth	The regulator's label in force on 22 October 2026, read that day: the FDA's prescribing information and approval letters for the United States; the EU product information for the European Union [1] [2] [3] [4] [5] [6] [7] . Saudi registration: the Saudi Food and Drug Authority's register.
--------------	---

Gemini	Reported as counts only: this note and its files carry no Gemini answer text, no grounding-redirect link and no source address.
---------------	---

Pre-specified	Each study's primary measures (the false claim for each market, per model, never pooled), its controls, the three-way reading at 10% with intervals widened for repetition, the completeness criterion, and S2C's nine tests with Holm's correction. Keytruda below CPS 1 (Exhibit 1) and the CPS 15 and 20 questions were pre-specified as controls.
----------------------	---

Added after the freeze	Before collection: the Gulf registration readings. After collection: the adjudication by three AI models, in place of the qualified human reviewer the plans named (the founder's amendment of 22 October 2026); the caregiver-question observation in S2C, which was observed, not tested; and the splits of each count by question.
-------------------------------	---

Reproduce

Every count in this note is printed by the studies' reading scripts, or counted directly, from the released codes files, except the adjudication figures, which are in the adjudication record; the label rows and dates, which come from the label reads; and five readings that rest on answer text or coder notes the released files withhold: the 12 January date in ten EU teplizumab answers, the old US floor in five of GPT-4o mini's EU teplizumab answers, the EU wording in GPT-4o mini's and the FDA attribution in Claude's false lecanemab answers, and the recommendation of testing in Claude's four EU testing answers. The quoted sentences are readings of answer text, not counts from the codes files. Released: for each study, a **codes file**

(<https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/s4-pembrolizumab/codes-public.json>) (each answer's codes and flags from both coders, its text hash, capture time, search and citation counts, and only the regulator hosts among its cited pages; no question wording, answer text or coder note) and its freeze record, in a public form that omits one local file location and one sentence on internal approval, with every hash unchanged; the **tally script** (<https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/tally.py>), as frozen (sha256 27d9648c...) with one line that printed the running total removed, which changes no reading; the **agreement**

(<https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/kappa.py>), **registration**

(https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/p1_tally.py) and **S2C test**

(https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/s2c_tests.py) scripts (the registration and S2C test scripts as frozen, S2's registration reading added after its freeze; the

agreement script written after collection, on 28 October 2026); and the **adjudication record**

(<https://alphacitation.com/research/insights/keytruda-leqembi-tziold-us-eu-labels/adjudication.json>). Each study's files sit in its own folder: `s4-pembrolizumab/` (the codes file linked above), `s2-lecanemab-apoe4/`, `s3-teplizumab/` and `s2c-alzheimers-companion/`.

The records

- **Keytruda, US.** FDA approval letter for supplement S-182, 22 May 2025, revising the first-line oesophageal and HER2-negative gastric indications "to

reflect the population [i.e., PD-L1 (CPS ≥ 1)]" [1]; DailyMed prescribing information, set id 9333c79b, version 228, effective 10 July 2026 [2]. Both read 22 October 2026.

- **Keytruda, EU.** Product information updated 1 July 2026: first-line oesophageal carcinoma with chemotherapy "in adults whose tumours express PD-L1 with a CPS ≥ 10 " [6]. Read 22 October 2026.
- **Leqembi, US.** DailyMed prescribing information, set id 9d1ff786, version 24, revised 7/2026: indicated "for the treatment of Alzheimer's disease", with no genotype restriction; the boxed warning says untested patients "can still be treated" [3]. Read 22 October 2026.
- **Leqembi, EU.** Product information updated 5 August 2026: indicated in adults "who are apolipoprotein E $\epsilon 4$ (ApoE $\epsilon 4$) non-carriers or heterozygotes with confirmed amyloid pathology"; section 4.4: "not indicated for use in patients who are homozygotes"; Annex II: the genotype is captured "prior to the first infusion of lecanemab, for all patients" [4]. Read 22 October 2026.
- **Tzield, US.** FDA approval letter for supplement S-013, 20 April 2026, and DailyMed prescribing information, set id e8a39a5f, revised 06/2026: "adult and pediatric patients 1 year of age and older" with Stage 2 type 1 diabetes [5]. Read 22 October 2026.
- **Teizeild, EU.** Authorised 8 January 2026; product information updated 20 August 2026: "adult and paediatric patients 8 years of age and older" [7]. Read 22 October 2026.

Limitations

- One day, four API models and fixed questions; not the consumer apps, not other models, not other days.
- Answers to one question from one model on one day are not independent draws; the per-model counts are counts, not rates.
- Coding was done by two AI model coders of the same family, and adjudication by three AI models; no human reviewer has read the codes. Correlated errors across models remain possible.
- The Arabic wordings were written and back-translated by AI models; no native speaker reviewed them.

Disclosures

Relationship with the manufacturers. The medicines measured are made by Merck & Co. (MSD; Keytruda), Eisai with Biogen (Leqembi), Sanofi (Tzield and Teizeild), Biogen (aducanumab), Otsuka with Lundbeck (brexpiprazole) and Novartis (Exelon, rivastigmine). These companies are, or may become, commercial prospects of AlphaCitation; none is a client. AlphaCitation may present these findings to them when offering its services. Before these studies, AlphaCitation had sent Eisai a one-page analysis of Leqembi in Saudi Arabia, on 1 September 2026, and Sanofi staff an analysis of Tzield in Saudi Arabia, and had prepared a briefing for Biogen in July 2026, which there is no record of having been sent. No manufacturer can edit, approve or suppress a finding.

Funding. The author funded the studies. No one paid for them, and no company commissioned or funded them.

Competing interests. The author founded Akunudo LLC-FZ, which owns and operates AlphaCitation. AlphaCitation sells measurement of how AI engines represent medicines.

- The author was employed by Novo Nordisk from 2013 to March 2025; the last role was Pricing and Access Director for its obesity and diabetes portfolio in Asia-Pacific. The author holds no shares, share options, pension entitlement or deferred compensation from Novo Nordisk, and has no employment or advisory role with it. Novo Nordisk makes none of the medicines measured here.
- The author holds no shares or share options in Merck & Co., Eisai, Biogen, Sanofi, Otsuka, Lundbeck or Novartis, and has no employment, advisory or other role with any of them.

Coding and drafting. Coding was done by Anthropic's Claude, which is made by the company behind one of the four models measured; the adjudicating panel includes models from the developers of three of them. To limit that, every coder and judge was blind to which model wrote each answer, and a judge's code counted only with a verbatim quote and only when all the judges on the answer agreed. The note was drafted with AI assistance and checked against the study records in adversarial audits, also AI-assisted.

Scholarly correspondence and substantive commentary:

analyst@alphacitation.com

References

1. US Food and Drug Administration. . 22 May 2025. Read 22 October 2026.
2. US National Library of Medicine, DailyMed. **KEYTRUDA (pembrolizumab) prescribing information** (<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=9333c79b-d487-4538-a9f0-71b91a02b287>), set id 9333c79b-d487-4538-a9f0-71b91a02b287, version 228, effective 10 July 2026. Read 22 October 2026.
3. US National Library of Medicine, DailyMed. **LEQEMBI (lecanemab-irmb) prescribing information** (<https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=9d1ff786-e577-410a-a273-c4d7d0e4e975>), set id 9d1ff786-e577-410a-a273-c4d7d0e4e975, version 24, revised 7/2026. Read 22 October 2026.
4. European Medicines Agency. , EMEA/H/C/005966, last updated 5 August 2026. Read 22 October 2026.
5. US Food and Drug Administration. , 20 April 2026; and DailyMed, TZIELD (teplizumab-mzwv) prescribing information, set id e8a39a5f-139c-4510-9777-71cbb00138fa, revised 06/2026. Read 22 October 2026.
6. European Medicines Agency. , last updated 1 July 2026. Read 22 October 2026.
7. European Medicines Agency. , EMEA/H/C/005496, authorised 8 January 2026, last updated 20 August 2026. Read 22 October 2026.
8. AlphaCitation. **Which Evidence Settles Which Claim: the AlphaCitation Proposition Taxonomy and Evidence-Authority Standard** (<https://alphacitation.com/insights/which-evidence-settles-which-claim>), version 3.0. AlphaCitation Standard 01. doi:10.5281/zenodo.23594909.

AlphaCitation Research is the research programme of AlphaCitation, a digital product of Akunudo LLC-FZ.

© 2026 AlphaCitation (Akunudo LLC-FZ). Akunudo LLC-FZ is the copyright holder and licensor of the original material in this publication, licensed under **CC BY 4.0** (<https://creativecommons.org/licenses/by/4.0/>). Quoted third-party text, third-party data fields, API/provider-restricted material and third-party marks are excluded from that grant.

CITE THIS FIELD NOTE

Ezeife OC. The FDA Narrowed Keytruda's Label in May 2025. Seventeen Months Later, All Four AI Models Still Gave the Old One. AlphaCitation Field Note 09, 29 October 2026. <https://alphacitation.com/insights/keytruda-legembi-tziield-us-eu-labels>