ANALÝZA A ZÁZNAM STAVU · s AI o AI

Vodoznak v textu od Claude: co měří, co neměří a koho se k srpnu 2026 vlastně týká

Pavel Horák · 16. srpna 2026 · stav ověřený k tomuto datu

Anthropic zavádí do textu generovaného modelem Claude neviditelný vodoznak a metodu poprvé pojmenoval 14. srpna 2026. Tento materiál rozebírá, co vodoznak technicky měří, ve kterých situacích z něj nezbude prakticky nic, co o jeho spolehlivosti říká nezávislý výzkum a proč se k datu publikace netýká žádného veřejně dostupného modelu Claude. Je psaný jako záznam stavu k datu, protože téma se bude rychle vyvíjet a starší zjištění se v něm nepřepisují, jen doplňují.

Anthropic zavádí do textu generovaného modelem Claude neviditelný vodoznak a metodu poprvé pojmenoval 14. srpna 2026. Tento materiál rozebírá, co vodoznak technicky měří, ve kterých situacích z něj nezbude prakticky nic, co o jeho spolehlivosti říká nezávislý výzkum a proč se k datu publikace netýká žádného veřejně dostupného modelu Claude. Je psaný jako záznam stavu k datu, protože téma se bude rychle vyvíjet a starší zjištění se v něm nepřepisují, jen doplňují. **Jak číst tento materiál.** Hlavní text zachycuje stav ověřený k 16. srpnu 2026 a zůstává neměnný. Pozdější vývoj se nepřepisuje do původních kapitol, ale přidává se datovaně do kapitoly 12. Pokud se některé kapitoly týká novější zjištění, je u ní uvedený odkaz na příslušný záznam. Přiložené PDF a zvuková verze odpovídají stavu k datu publikace a při aktualizacích se neregenerují.

1. Shrnutí stavu k 16. srpnu 2026

Anthropic 11. srpna 2026 aktualizoval nápovědní článek o značení obsahu generovaného modelem Claude a 14. srpna k němu vydal podrobné otázky a odpovědi, ve kterých poprvé pojmenoval použitou metodu. Body, které k datu publikace platí, jsou tyto.

- Jde o variantu přístupu SynthID-Text, který publikoval Google DeepMind v recenzovaném článku v časopise Nature v roce 2024. Metoda patří do rodiny přístupů navazujících na návrh Scotta Aaronsona z roku 2022.
- Do textu se nic nepřidává a nejsou v něm žádné skryté znaky. Mění se pouze zdroj náhody při výběru slova.
- Značení se neúčtuje v tokenech, negeneruje vyšší cenu a nezpomaluje generování.
- Vodoznak neobsahuje žádné identifikační údaje a podle Anthropicu ho nelze dohledat ke konkrétní osobě, organizaci ani konverzaci.
- Detekční rozhraní zatím neexistuje. Anthropic ho slibuje bez uvedení termínu.
- U souborů typu .png, .jpg nebo .svg jde o zcela jinou techniku, tedy o podepsaná metadata podle otevřeného standardu C2PA. To není vodoznak, ale připojená poznámka, kterou lze běžným zpracováním souboru odstranit.
- Značení se týká modelů uvedených 2. srpna 2026 a později. Poslední veřejně uvedený model Claude je z 24. července 2026. Z toho plyne závěr rozebraný v kapitole 5.

- Systémy uvedené na trh před 2. srpnem 2026 mají přechodné období na doplnění značení do 2. prosince 2026. Tohle je nejbližší tvrdý termín celého tématu.

2. Časová osa

Pro pozdější porovnání je užitečné mít události seřazené s daty, protože se v prvním týdnu překrývaly a část zpravodajství je slučuje do jedné.

- **Červenec 2026.** Anthropic podepisuje spolu s přibližně 190 dalšími signatáři evropský kodex pro transparentnost obsahu generovaného umělou inteligencí, navázaný na článek 50 odst. 2 nařízení AI Act.
- **2. srpna 2026.** Datum, od kterého se odvíjejí povinnosti značení. Modely uvedené tímto dnem a později mají značení podporovat od startu.
- **11. srpna 2026.** Anthropic aktualizuje nápovědní článek o značení obsahu. Stránka sama přesné datum revize neuvádí, je na ní pouze údaj o aktualizaci v daném týdnu. To je pro pozdější ověřování slabina zdroje a je nutné s ní počítat.
- **12. srpna 2026.** Podle sekundárních zdrojů se veřejně vyjadřuje inženýr Anthropicu a potvrzuje tři věci, tedy že detekční rozhraní bude použitelné i pro běžné uživatele, že model sám o svém značení neví, a že podobné značení připravují i další laboratoře. Toto je jediná položka časové osy, kterou se mi nepodařilo doložit primárním zdrojem.
- **14. srpna 2026.** Anthropic publikuje podrobné otázky a odpovědi k fungování vodoznaku a poprvé pojmenovává metodu.
- **16. srpna 2026.** Datum ověření všech tvrzení v tomto materiálu proti primárním zdrojům.

3. Jak vodoznak funguje

Jazykový model generuje text slovo po slově a u velké části rozhodnutí má několik prakticky rovnocenných možností. Ve větě „počasí bylo dnes chladné a…" nezáleží na tom, jestli přijde „zataženo" nebo „šedivo", význam zůstává stejný. O výběru mezi rovnocennými kandidáty běžně rozhoduje náhodné číslo.

Vodoznak nepřidává do textu nic navíc. Mění pouze zdroj té náhody. Místo generátoru náhodných čísel se použije tajný klíč a několik předchozích slov. Volba zůstává z pohledu čtenáře náhodná a nepoznatelná, ale kdo má klíč, může zpětně spočítat pravděpodobnost, že sekvenci vybíral model pracující s tímto klíčem.

Technicky se to řeší úpravou vzorkovací procedury. Kandidáti na další token vstupují do struktury, které se říká turnajové vzorkování, a o postupujícím rozhoduje funkce odvozená z tajného klíče a předchozího kontextu. Metoda je navržená jako nedistorzní, tedy tak, aby se výsledné rozdělení pravděpodobností v průměru nelišilo od nevodoznakovaného modelu, a je slučitelná se spekulativním vzorkováním, což je podmínka nasaditelnosti v produkčním provozu bez znatelného zpomalení. Právě z nedistorzní povahy plyne tvrzení, že se nemění kvalita výstupu.

Anthropic k tomu používá přirovnání k deskové hře, ve které místo házení kostkou hráči postupují podle číslic čísla pí. Hra probíhá stejně a hráčům je to jedno, ale kdo zná pí, pozná zpětně, že se hrálo podle něj. Přirovnání je přesné v tom podstatném, tedy že se nemění pravidla hry, ale zdroj náhodnosti.

Důležitý důsledek, který se v české publicistice ztrácí: vodoznak není skrytý znak, neviditelná mezera ani metadatová značka. Není to nic, co by šlo z textu vymazat hledáním a nahrazením. Je to statistická vlastnost výběru slov napříč celým textem. Proto také přežije zkopírování a vložení

jinam a proto zároveň slábne všude tam, kde model slova nevybírá.

U souborů, tedy u obrázků a vektorové grafiky, jde o úplně jinou techniku. Připojí se kryptograficky podepsaná poznámka v metadatech podle standardu C2PA. Ta říká, že soubor byl vytvořen nebo zpracován modelem, a umožňuje poznat pozdější manipulaci. Na rozdíl od textového vodoznaku ale zmizí při běžném zpracování, tedy při převodu formátu, opětovném uložení, optimalizaci nebo pořízení snímku obrazovky. Anthropic to sám uvádí mezi omezeními. Slučovat obě techniky pod jedno slovo je věcná chyba.

4. Co vodoznak měří a co neměří

Tohle je jádro celé věci a zároveň místo, kde se veřejná debata míjí s technickou realitou.

Vodoznak umí odpovědět na jedinou otázku, totiž jaká je pravděpodobnost, že se na daném textu podílel Claude. Neodpovídá na otázku, jestli text psal člověk. Nepozná text z jiného modelu, protože ten má buď jiný klíč, nebo úplně jinou metodu. A podle vlastní formulace Anthropicu nedokáže odlišit situaci „Claude to napsal" od situace „Claude to výrazně upravil".

Z toho plynou dva samostatné a stejně důležité závěry, které se v diskuzi obvykle zaměňují.

Nález značky není důkaz autorství. Znamená, že se textu pravděpodobně dotkl model. Neříká nic o tom, kdo přišel s myšlenkou, kdo udělal rešerši, kdo napsal první verzi a kolik toho člověk změnil. Text, který si autor napsal sám a nechal si opravit interpunkci, může nést značku. Text, který za autora vymyslel a napsal model, ji nést nemusí, pokud prošel důkladným přepisem.

Absence značky není důkaz lidského autorství. Anthropic sám vyjmenovává situace, ve kterých značka chybí, přestože obsah vznikl s pomocí modelu. Patří sem starší model bez podpory značení, silná editace nebo parafráze, příliš krátká pasáž, odstraněná metadata u souborů a nepodporovaná kombinace platformy nebo typu souboru.

Vodoznak je tedy nástroj na zjišťování provenience, nikoli na určování autorství. To jsou dvě různé otázky a zaměňují se proto, že veřejnost i školy potřebují odpověď na tu druhou, zatímco technologie odpovídá na tu první.

5. Koho se značení k datu publikace vlastně týká

Tady je zjištění, které v české ani v zahraniční publicistice nemá pokrytí odpovídající jeho významu, a proto ho uvádím i s postupem, kterým jsem k němu došel.

Vycházím ze dvou doložených premis.

První premisa. Anthropic v nápovědním článku i v odpovědích uvádí, že modely Claude uvedené 2. srpna 2026 a později podporují značení od startu. U modelů uvedených před tímto datem odkazuje na přechodné období podle zákona a uvádí, že na doplnění značení se teprve pracuje a bude se zavádět v následujících měsících. Text odpovědí navíc hned v první větě mluví o budoucích modelech Claude.

Druhá premisa. Poslední veřejně uvedený model Claude je Claude Opus 5 z 24. července 2026, tedy devět dní před rozhodným datem. Ostatní modely v aktuální nabídce jsou starší. Claude Sonnet 5 byl uveden 30. června 2026, Claude Fable 5 spolu s modelem Mythos 5 dne 9. června 2026 a Claude Haiku 4.5 v říjnu 2025. Žádný veřejně dostupný model tedy nespadá do kategorie uvedené 2. srpna 2026 a později.

Závěr. Z obou premis plyne, že k 16. srpnu 2026 nevodoznakuje text žádný veřejně dostupný model Claude. Vlna rušení předplatného, která se v prvním týdnu objevila, byla tedy reakcí na vlastnost, kterou model používaný stěžovateli v tu chvíli neměl.

Do kdy to může platit. Formulace Anthropicu o následujících měsících vypadá jako neurčitá, ve skutečnosti ale neurčitá není. Pro povinnost strojově čitelného značení podle článku 50 odst. 2 platí omezené přechodné období a generativní systémy uvedené na trh před 2. srpnem 2026 mají na doplnění značení čas do 2. prosince 2026. Toto přechodné období se netýká ostatních povinností transparentnosti, ty platí od srpna bez odkladu. Prosincové datum je tedy tvrdý regulatorní horizont, do kterého musí být doznačeny i modely, kterých se dnes značení netýká. Tato lhůta pochází z omnibusové novely nařízení, nikoli z kodexu, což se v komentářích občas zaměňuje.

Je nutné zdůraznit, čím tento závěr je a čím není. Není to tvrzení Anthropicu, firma to takto nikde neformulovala a pravděpodobně ani neformuluje. Je to odvození ze dvou samostatně doložených faktů. Slabinou odvození je, že veřejný seznam modelů nemusí zachycovat tiché aktualizace, a druhou slabinou je, že přesné datum poslední revize nápovědního článku není uvedeno. Zároveň je to zjištění s nejkratší trvanlivostí v celém materiálu. První nový model po rozhodném datu ho zneplatní a v tu chvíli se to promítne do kapitoly 12.

6. Čtyři situace a kolik signálu v nich zbude

Anthropic sám v odpovědích popisuje, kde metoda funguje dobře a kde skoro vůbec. Následující přehled vychází z jeho vlastních formulací, protože právě tohle je pro běžného uživatele nejpraktičtější část celé věci.

Situace	Kolik slov volí model	Síla značky
Psaní delšího textu od nuly	všechna	nejvyšší, roste s délkou
Překlad	všechna	plná, myšlenky jsou cizí, ale slova volí model
Korektura cizího textu	jednotky oprav	velmi nízká, často neregistrovatelná
Kód	málo, výstup musí být přesný	výrazně slabší, uplatní se hlavně v komentářích
Faktický a odborný text	omezeně, správná odpověď bývá jediná	řidší než u volného textu
Krátká pasáž	málo rozhodnutí celkem	nespolehlivá

Logika je ve všech řádcích stejná. Značka žije jen tam, kde má model svobodu volby mezi rovnocennými možnostmi. Tam, kde je správná odpověď jediná, se značka neuplatní, protože jinak by ji zaplatila věcná správnost. Anthropic to ilustruje na příkladu, kdy po slovech o Newtonově nejslavnějším díle existuje jen jedno správné pokračování, a na příkladu součtu dvou a dvou.

Pro dvě skupiny uživatelů, které si nejvíc stěžovaly, z toho plyne poměrně jasná odpověď. Kdo si nechává kontrolovat pravopis vlastního textu, u toho nemusí zbýt nic, na čem by se dala značka postavit. A u kódu je značení výrazně slabší už z principu, protože kód musí být přesný.

Opačný pól je překlad. Ten nese značku v plné síle, i když myšlenky ani formulace nejsou modelu vlastní. Právě překlad je nejnázornější ukázkou toho, proč se provenience a autorství nedají zaměňovat.

7. Co o spolehlivosti říká nezávislý výzkum

Anthropic uvádí omezení poctivěji, než bývá u firemní komunikace obvyklé. Akademická literatura je přesto skeptičtější, a to v rovině, která má přímý dopad na to, jestli je takový nález použitelný jako podklad pro obvinění.

Hodnocení z roku 2025 ukázalo, že metoda SynthID-Text je zranitelná vůči úpravám zachovávajícím význam, tedy vůči parafrázi, přesouvání částí textu a zpětnému překladu, které detekovatelnost výrazně snižují. Teoretická analýza z března 2026 na to navázala rozborem samotné skórovací funkce a ukázala, že určité konfigurace jsou proti odstranění značky podstatně slabší než jiné.

Nejvíc vypovídající je ale práce z července 2026, která se ptá jinak než ostatní. Neptá se, jak dobře metoda funguje jako klasifikátor, ale jestli obstojí jako forenzní důkaz. Navrhuje k tomu hodnoticí rámec a její výsledky jsou tvrdé.

- Míra falešně negativních výsledků u testované konfigurace SynthID byla vysoká ještě před jakýmkoli zásahem do textu.
- Po parafrázi zmizela značka téměř ve všech případech, kdy byla původně nalezena.
- Malá, ale nikoli zanedbatelná část parafrázovaných textů psaných člověkem byla označena jako vygenerovaná umělou inteligencí.
- Velká část nedotčeného vygenerovaného textu skončila v pásmu nejistoty, tedy bez jednoznačného výsledku v obou směrech.
- Žádná ze tří posuzovaných metod nesplnila více než dva z pěti Daubertových faktorů, tedy kritérií přípustnosti znaleckého důkazu v americkém soudnictví.

Práce zároveň cituje starší srovnávací benchmark, ve kterém úspěšnost detekce u SynthID klesla z hodnoty blízké jistotě zhruba na polovinu už při mírné parafrázi, a to bez použití specializovaných nástrojů.

Tato kapitola nemá být návodem a záměrně neuvádí konkrétní parametry ani postupy. Její smysl je jiný a dá se shrnout do jedné věty. Falešně pozitivní nález je pro běžného uživatele větší riziko než falešně negativní. Riziko neleží v tom, že někoho vodoznak odhalí, ale v tom, že se z pravděpodobnostního signálu udělá rozhodnutí, které se tváří jako jistota, a doplatí na to člověk, který model nepoužil.

8. Co po publikujících vyžaduje AI Act

Regulační pozadí je důvod, proč se to celé děje, a zároveň místo, kde se nejčastěji chybuje ve výkladu.

Od 2. srpna 2026 se uplatňuje článek 50 nařízení AI Act, který ukládá povinnosti transparentnosti ve čtyřech oblastech, tedy u přímé interakce s člověkem, u obsahu generovaného umělou inteligencí, u rozpoznávání emocí a biometrické kategorizace a u deepfaků a textů k věcem veřejného zájmu. Za jejich nedodržení hrozí podle sankčního rámce nařízení pokuty až 15 milionů eur nebo tři procenta celosvětového ročního obratu, podle toho, která částka je vyšší. Evropská komise k výkladu článku 50 zveřejnila 20. července 2026 finální pokyny, které nahradily květnový konzultační návrh a slouží jako hlavní referenční dokument pro dozorové orgány.

Dvě data v tom hrají různou roli a pletou se. Povinnosti podle článku 50 platí od 2. srpna 2026 bez ohledu na to, kdy byl systém uveden na trh. Výjimkou je jediná povinnost, totiž strojově čitelné značení podle odstavce 2, u které mají systémy uvedené na trh před tímto datem přechodné období do 2. prosince 2026.

Pro publikující je podstatné ještě jedno pravidlo, které se v českém zpravodajství nezmiňuje vůbec. Obsah, který byl vygenerován a zpřístupněn před 2. srpnem 2026, se zpětně označovat nemusí. Pokud ale obsah vznikl před tímto datem a publikuje se až po něm, povinnost označení se na něj vztahuje. Kdo má rozpracované texty ve frontě, měl by si tohle ohlídat. Anthropic zavádí značení celosvětově, protože podle vlastního vyjádření zatím nemá spolehlivý způsob, jak ho omezit jen na evropský trh, a uvádí, že bude různé přístupy dál vyhodnocovat.

Podstatný je ale výklad povinnosti na straně toho, kdo obsah publikuje, a ten se od povinnosti poskytovatele modelu liší. U textů zveřejněných za účelem informování veřejnosti o věcech veřejného zájmu platí povinnost zřetelného označení tehdy, pokud text neprošel lidskou revizí nebo redakční kontrolou a pokud za něj někdo nepřevzal redakční odpovědnost.

Jinými slovy, zákon u publikovaného textu neřeší, kdo ho napsal, ale kdo za jeho výslednou podobu ručí. Rozhodující je skutečná redakce, nikoli formální. To je zásadní posun oproti tomu, jak se o věci mluví ve veřejné debatě, kde se pořád hledá odpověď na otázku autorství.

Poslední detail, který stojí za pozornost, pochází z právních komentářů ke kodexu a nedokázal jsem ho ověřit přímo v jeho textu, proto ho uvádím s touto výhradou. Kodex podle nich uznává, že žádná jednotlivá technika značení dnes nesplňuje všechny požadavky současně, a proto počítá s vícevrstvým přístupem, tedy s nejméně dvěma vrstvami strojově čitelného značení tam, kde je to potřeba. Pokud to platí, je vodoznak zamýšlený jako první vrstva z několika, ne jako konečné řešení.

9. Co z toho plyne pro toho, kdo píše a publikuje

Praktické závěry jsou čtyři a všechny míří spíš k procesu než k technologii.

Rušit předplatné kvůli vodoznaku je předčasná reakce. K datu publikace se značení netýká žádného dostupného modelu, a až se dotýkat bude, u korektury vlastního textu z něj zůstane minimum. To se ale změní a je rozumné počítat s tím, že u delších generovaných textů značka jednou bude.

Nález značky nesmí sám o sobě nést obvinění. Platí to pro školy, pro redakce i pro zadavatele, kteří kontrolují dodané texty. Anthropic sám uvádí, že nález není plně průkazný, a nezávislý výzkum ukazuje, že falešně pozitivní výsledky u parafrázovaných lidských textů nejsou nulové. Nález je důvod k rozhovoru, ne k rozhodnutí.

Absence značky nic nedokazuje. Kdo si od dodavatelů nechává potvrzovat, že text nevznikl s pomocí umělé inteligence, opírá se o test, který v obou směrech selhává. Nulový nález u starého modelu, krátkého textu nebo přepsané pasáže vypadá stejně jako nulový nález u textu psaného člověkem.

Redakční politika je potřebnější než detekce. Kdo publikuje, měl by mít napsané, jaká míra zapojení modelu spouští uvedení u textu. Věty „Claude to napsal" a „Claude to zkorigoval" popisují dvě různé situace a rozdíl mezi nimi musí být pojmenovaný předem, ne až se někdo zeptá. Zákon se ptá po redakční odpovědnosti a tu nelze doložit zpětně technickým testem, jen předem nastaveným procesem.

10. Poznámka k tomuto materiálu

Tento text vznikl ve spolupráci s modelem Claude a web, na kterém je publikován, to uvádí u každé studie. Za výběr úhlu, ověření zdrojů, formulaci závěrů i za případné chyby ručí podepsaný autor. To je přesně ta poloha, kterou článek 50 po publikujících chce, tedy pojmenovaná lidská

odpovědnost, nikoli technický otisk v textu. Uvádím to proto, že by bylo nedůsledné psát o redakční odpovědnosti a zároveň ji u vlastního materiálu nepřiznat.

Součástí přípravy byl i nezávislý audit textu provedený konkurenčním modelem, konkrétně Google Gemini Pro, se zadáním zkontrolovat všechna data a navrhnout doplnění. Uvádím ho, protože jeho výsledek ilustruje rozdíl mezi redakční kontrolou a jejím zdáním.

Audit přinesl jeden údaj, který mi unikl a který studii podstatně zlepšil, totiž existenci přechodného období do 2. prosince 2026. Ověřil jsem ho u několika na sobě nezávislých právních zdrojů a je v kapitolách 1, 5 a 8. Zároveň ale audit obsahoval tvrzení, která do textu nešla převzít. Popis jednoho z akademických útoků na vodoznak byl věcně chybný, protože zaměňoval turnajové vrstvy vodoznakovacího schématu za skryté vrstvy neuronové sítě a spojoval je s parametrem teploty, což citovaná práce netvrdí. Rozbor Daubertových faktorů po jednotlivých položkách nebyl obsahem citované práce, ale konstrukcí auditujícího modelu. Několik technických parametrů modelu a jeden odborně znějící termín se nepodařilo dohledat vůbec.

Praktický závěr je ten, že audit byl užitečný, ale jen proto, že prošel druhou kontrolou. Nezkontrolovaný by do textu o spolehlivosti detekce vnesl nedoložená tvrzení, tedy přesně to, před čím tento materiál varuje. Podrobnější rozbor toho, jak vypadá selhání dlouhého generovaného výstupu, je v samostatném materiálu Anatomie rozpadu výstupu jazykového modelu z června 2026.

11. Otevřené otázky

Následující otázky jsou položené k 16. srpnu 2026 a zůstávají v tomto znění i poté, co se zodpoví. Odpovědi se doplňují do kapitoly 12.

- Kdy bude zveřejněno detekční rozhraní a v jaké podobě, tedy jestli bude dostupné i mimo firemní zákazníky.
- Stihne Anthropic doznačit modely uvedené před 2. srpnem 2026 do prosincové lhůty a půjde o dodatečné doznačení stávajících modelů, nebo jen o nové verze.
- Hypotéza k ověření, nikoli zjištění: kromě regulatorního důvodu má plošné značení i vedlejší užitek pro samotného poskytovatele, protože držitel klíče dokáže ve stažených webových datech rozpoznat a vyfiltrovat vlastní starší výstupy před tréninkem dalšího modelu. Zda tuto motivaci někdo z poskytovatelů přizná, je otevřené.
- Který model bude prvním veřejně uvedeným modelem Claude se značením od startu.
- Zveřejní Anthropic míru falešně pozitivních a falešně negativních výsledků své konkrétní konfigurace, nebo zůstane u obecného popisu.
- Vydá OpenAI textový vodoznak, když u obrázků a zvuku značení už používá.
- Objeví se druhá vrstva strojově čitelného značení, kterou předpokládají právní komentáře ke kodexu.
- Padne první doložený případ, ve kterém detekční nález povede k obvinění, a jak s ním naloží škola, redakce nebo soud.

12. Aktualizace

Tato kapitola je při publikaci prázdná. Aktualizace se sem přidávají datovaně, nejnovější nahoře, a mají tuto podobu: datum, dotčené kapitoly, popis změny a zdroj. Rozsáhlejší posun se zaznamenává jako nové shrnutí stavu k novému datu, které nenahrazuje shrnutí v kapitole 1, ale staví se vedle něj. Původní text kapitol se nikdy nepřepisuje ani nemaže, protože jeho hodnota spočívá právě v tom, že zachycuje, co se vědělo v daný den. Do dotčené kapitoly se doplňuje

pouze odkaz na příslušný záznam, aby čtenář neodešel se zastaralou informací.

Přiložené PDF a zvuková verze odpovídají stavu k 16. srpnu 2026 a při aktualizacích se neregenerují. Platným zdrojem aktuálního stavu je vždy tato stránka.

Zdroje

Všechny odkazy byly ověřeny 16. srpna 2026.

1 Anthropic: How Claude's text watermark works, publikováno 14. srpna 2026. anthropic.com

2 Anthropic: How Claude marks AI-generated content, nápovědní článek, bez uvedeného přesného data revize. support.claude.com

3 Dathathri a kolektiv: Scalable watermarking for identifying large language model outputs, Nature, 2024. Původní publikace metody SynthID-Text.

4 Robustness Assessment and Enhancement of Text Watermarking for Google's SynthID, arXiv 2508.20228, 2025. Hodnocení odolnosti vůči úpravám zachovávajícím význam.

5 On Google's SynthID-Text LLM Watermarking System, arXiv 2603.03410, březen 2026. Teoretická analýza skórovacích funkcí.

6 AI Watermark Evidence Fails Forensic Readiness: An Empirical Evaluation, arXiv 2607.16010, červenec 2026. Zdroj údajů o falešně pozitivních a falešně negativních výsledcích a o Daubertových faktorech.

7 Evropská komise: Code of Practice on Transparency of AI-Generated Content, červenec 2026, přibližně 190 signatářů.

8 Nařízení Evropského parlamentu a Rady o umělé inteligenci, článek 50 a sankční ustanovení.

9 Evropská komise: finální pokyny k povinnostem transparentnosti podle článku 50, zveřejněné 20. července 2026.

10 Právní komentáře k přechodnému období do 2. prosince 2026 pro systémy uvedené na trh před 2. srpnem 2026, ověřeno u několika na sobě nezávislých zdrojů, mimo jiné Cooley, Stibbe, Bird & Bird a Faegre Drinker, a u informační stránky Evropské komise ke kodexu.

11 Sekundární zdroje k časové ose a k výkladu povinností publikujících, zejména TechCrunch, SiliconANGLE a české zpravodajství. Položka časové osy k 12. srpnu 2026 a tvrzení o vícevrstvém značení pocházejí výhradně ze sekundárních zdrojů a jsou v textu takto označeny.

08 2026 Vytvořil Pavel Horák s Claude OPUS - s AI o AI (sAIoAI.cz) | Vibe Coding prakticky | RAG, co to je?. Návrat na sAIoAI