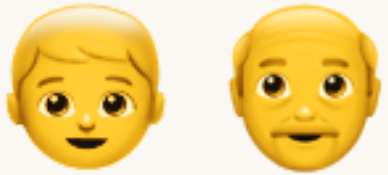


～この画像、信じていいの?～

博士とケンタの会話で学ぶ C2PA

この写真、信じていいの?

📱 ケンタがスマホで1枚の写真を見ている

👦 はかせ、この写真さ……本物なの? だれかが作った偽物かもしれないじゃん。信じていいの?

👴 ほっほ。それがいちばん大事な問いじゃ。……正直に言うとな、いまのインターネットは、その「本物?」にほとんど答えられんのじゃよ。

👦 えっ、答えられないの? じゃあみんな、何を見て信じてるの?

👴 「なんとなく、それっぽいから」じゃな。写真そのものには、だれが撮ったかも、いつ撮ったかも、どこにも書いておらん。手がかりが、ほぼ無いのじゃ。

👦 手がかりなしで信じるの、こわいな……。

👴 こわいじゃろ。だからこそ、この問いから始めるのじゃ。「どうすれば、ちゃんと根拠を持って信じられるか」——な。

この写真、信じていいの？

📱 ケンタがスマホで1枚の写真を見ている

👦 はかせ、この写真さ……本物なの？ だれかが作った偽物かもしれないじゃん。信じていいの？

📌 **参考**

強力な作成・編集技術の普及により、メディアの来歴（provenance）を確立することが、透明性と信頼の確保に不可欠になっている。

"With the increasing velocity of digital content and the increasing availability of powerful creation and editing techniques, establishing the provenance of media is critical to ensure transparency, understanding, and ultimately, trust."

C2PA Technical Specification 2.4 §1.1 — https://spec.c2pa.org/specifications/specifications/2.4/specs/C2PA_Specification.html

👦 手がかりなしで信じるの、こわいな……。

👴 こわいじゃろ。だからこそ、この問いから始めるのじゃ。「どうすれば、ちゃんと根拠を持って信じられるか」――な。

AI に見抜いてもらえばいいんじゃない?

わかった! AI に「これ偽物?」って聞けばいいじゃん! AI、なんでもできるんでしょ?

その手はある。じゃがな、AI は写真の中身を見て「たぶん偽物」と“推測”しているだけで、100% ではないのじゃ。

推測かあ。当たらないこともあるってこと?

そうじゃ。しかも、偽物を“作る”AI が上手くなれば、“見抜く”AI はすり抜けられる。いたちごっこで、どこまでいっても「たぶん」止まりなのじゃ。

「たぶん本物」じゃ、安心できないなあ……。

うむ。後から AI で見抜くのには、どうしても限界があるのじゃ。

TechThanks

AI に見抜いてもらえばいいんじゃない?

わかった! AI に「これ偽物?」って聞けばいいじゃん! AI、なんでもできるんでしょ?

その手はある。じゃがな、AI は写真の中身を見て「たぶん偽物」と"推測"しているだけで、100% ではないの

📌 **参考**

来歴（C2PA）は検出を置き換えるのではなく補完する。

"It complements media literacy, fact-checking, and digital forensics approaches such as deep-fake detection by providing an infrastructure to record all of that information in a tamper-evident structure, representing the provenance of any asset."

C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

補足:「検出は確率的（100% でない）」という点は一般的な整理で、C2PA 仕様自体の主張ではない。

うむ。後から AI で見抜くのには、どうしても限界があるのじゃ。

そもそも、なんで今になって問題なの？

でもさ、「本物か分からない」なんて昔からあったんじゃないの？ なんで今なの？

昔は、本物そっくりの偽物を作るのが、とても大変だったからじゃ。専門の技術と手間と時間が要った。「世の中の写真は、だいたい本物だろう」――その暗黙の前提で、世界は回っておった。

あー。わざわざ偽物を作る人が、そもそも少なかったんだ。

ところが生成 AI で、その前提が崩れた。だれでも、一瞬で、本物と見分けのつかない画像や声を作れる。偽物を作るコストが、ほぼゼロになったのじゃ。

困ってなかったんじゃなくて……困る前提が、最近できたんだ。

そうじゃ。だから「これは本物か」を、勘でも後追いでもなく、確かな“仕組み”で確かめる必要が出てきたのじゃ。

そもそも、なんで今になって問題なの？

でもさ、「本物か分からない」なんて昔からあったんじゃないの？ なんで今なの？

昔は、本物そっくりの偽物を作るのが、とても大変だったからじゃ。専門の技術と手間と時間が要った。「世

📌 参考

生成 AI で作られた・編集されたコンテンツを識別したい、という関心の高まりが背景にある。

"People are increasingly concerned about being able to identify content that has been generated or edited by generative AI systems, or conversely, content that is generally unadulterated since its capture by e.g. a camera."

C2PA FAQ — https://c2pa.org/faqs/

補足: 「偽造コストの激変」という歴史的経緯は一般的な整理であり、C2PA 仕様自体の主張ではない。

そうじゃ。だから「これは本物か」を、勘でも後追いでもなく、確かな“仕組み”で確かめる必要が出てきたのじゃ。

中身じゃなく「来歴」を見る、ってどういうこと?

そこで発想を変える。中身が本物かを当てるのは、もうやめる。かわりに「だれが・いつ・何で作ったか」という“履歴”を、作った瞬間から、はんこ（署名）付きで写真にくっつけておくのじゃ。

サインなんて、後から書き換えられるんじゃない?

そこがミソでな。この“はんこ”は、写真の中身と暗号的に結びついておる。あとから中身を1ドットでもいじると、はんこが壊れて「改ざんされました」とすぐ分かるのじゃ。

お――! いじったらバレるんだ。

この“はんこ付きの来歴”そのものを、Content Credentials（コンテンツ・クレデンシャル）という。そして、その作り方を世界共通で決めた標準が C2PA じゃ。

Content Credentials が画像に“付くもの”で、C2PA がその“ルール”ってことか。

中身じゃなく「来歴」を見る、ってどういうこと?

📌 **参考**

 CR ピン（コンテンツ上に表示されるマーク）

 content credentials ロゴ

改変は暗号的な結び付きを壊し、改ざんとして検知される（tamper-evident）。

"Any modification—intentional or accidental—will break this cryptographic linkage, signalling tampering."

C2PA FAQ — https://c2pa.org/faqs/

Content Credentials（付くモノ）の仕様を、C2PA（団体・標準）がホストする。

"The Content Credentials specification is a project hosted by the Coalition for Content Provenance and Authenticity (C2PA)."

Content Credentials 公式 — https://contentcredentials.org/ ／ アイコン出典 — github.com/contentauth/verify-site（CR ピン）・contentcredentials.org（ロゴ）

それ、ふつうの EXIF じゃダメなの?

👧 写真ってもともと「撮った日」とか「カメラの種類」とか入ってるよね。EXIF ってやつ。あれじゃダメなの?

👴 EXIF では足りん。EXIF は、だれでも自由に書き換えられるし、消せる。撮影日も場所も、差し替え放題じゃ。

👧 あー、たしかに。日付とか、後から変えられそう。

👴 しかも"はんこ"が無いから、書き換えられても気づけんし、「だれが書いたのか」も分からん。いわば"鉛筆書きのメモ"なのじゃ。

👧 鉛筆書き。消しゴムで消せちゃう。

👴 Content Credentials は、同じ「いつ・だれが・何で」を書くのでも、署名入り・改ざんすると壊れる形で書く。鉛筆メモと、公証人のはんこ付き証明書くらい違うのじゃ。

それ、ふつうの EXIF じゃダメなの?

写真ってもともと「撮った日」とか「カメラの種類」とか入ってるよね。EXIF ってやつ。あれじゃダメなの?

EXIF では足りん。EXIF は、だれでも自由に書き換えられるし、消せる。撮影日も場所も、差し替え放題じ

📌 参考

C2PA は EXIF/XMP/IPTC などの既存メタデータを署名付き assertion として包み、改ざん検知可能・暗号的に検証可能にする。

"The C2PA model is designed to interoperate with standard metadata formats like IPTC, XMP, and EXIF. It can encapsulate these metadata types as assertions within a Content Credential, making them tamper-evident and cryptographically verifiable."

C2PA FAQ — https://c2pa.org/faqs/

Content Credentials は、同じ「いつ・だれが・何で」を書くのでも、署名入り・改ざんすると壊れる形で書く。鉛筆メモと、公証人のはんこ付き証明書くらい違うのじゃ。

でも、来歴なんてウソでも書けるよね?

👦 あ、でもさ。その"履歴"自体、ウソを書いて貼ればいいんじゃないの?

……それは本当に鋭い。そのとおり、来歴はだれでも付けられる。「来歴がある＝信頼できる」ではないのじゃ。 👴

👦 えー! じゃあ意味ないじゃん!

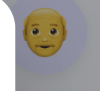
ここが肝心でな。署名が保証するのは「この主張をしたのは、確かにこの署名者だ」ということじゃ。 👴

ウソを書けば、「そのウソを主張した責任」が、署名者に永久に貼りつく。逃げられんのじゃ。 👴

👦 正直さは強制できないけど、無責任な言いっぱなしを"割に合わなく"するんだね。

でも、来歴なんてウソでも書けるよね?

あ、でもさ。その"履歴"自体、ウソを書いて貼ればいいんじゃないの?



📌 参考

Content Credentials が付いているという一点だけで信頼してはならない。

"no assumption should be made about the trustworthiness of a particular asset purely based on its usage of Content Credentials"

C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

署名者に帰属するのは、署名者が自ら作った主張（created_assertions）のみ。

"Only the set of assertions that are listed in the created_assertions field of the Content Credential's Claim are attributed to the signer."

C2PA Explainer 2.4 — 同上

だれが確かめるの？“信頼できる署名者”はだれが決めるの？

その来歴って、だれが読んで確かめるの？あと「信頼できる署名者」って、だれが決めるの？

読む係が Verifier（検証者）じゃ。スマホの表示機能や検証ツールがこれで、①来歴が改ざんされていないか、②署名者が“信頼できる名簿”に載った保証人（CA）につながるか、を確かめる。

名簿？

うむ。CA（認証局）は「この鍵の持ち主は確かにこの会社だ」と身元を保証する係。そして C2PA が、信頼できる CA の名簿――トラストリスト――を公開しておる。ブラウザの鍵マーク（SSL/TLS）と同じ、実証済みの仕組みじゃ。

勝手に「おれが保証人だ」って名乗るのは、ダメなんだね。

ダメじゃ。名簿は公開されていて、だれでも確認できる。「C2PA 対応です」とただ名乗るだけの製品と、審査を通った適合製品も、はっきり区別されるのじゃ。

だれが確かめるの？“信頼できる署名者”はだれが決めるの？

その来歴って、だれが読んで確かめるの？あと「信頼できる署名者」って、だれが決めるの？

📌 参考

Validator は Manifest の整合性を検証し、署名者が既知のトラストリストに関連づくかを確認する。トラストモデルは SSL/TLS や PDF 署名と同じ技術が土台。

"This trust model is based on the same technology behind SSL/TLS and PDF signatures and can be combined with established identity frameworks..."

C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

CA は公開鍵を身元に結びつける証明書を発行・署名・失効させる。Trust List は C2PA が管理する X.509 トラストアンカーのリスト。

"A trusted entity that issues, signs, and revokes digital certificates that bind public keys to subscriber identities."

C2PA Conformance Program v0.1 — https://github.com/c2pa-org/conformance-public

査を通った適合製品も、はっきり区別されるのじゃ。

で、結局この写真は「本物」なの？

ここまでやったら、この写真は「本物」だって分かるの？

——いや。ここが、いちばん大事なところじゃ。Content Credentials は「中身が本物だ」とは、絶対に言わん。

ええっ!? ここまでやって!?

言えるのは「信頼できるだれかが、改ざんされていない来歴を保証している」まで。緑のチェックの意味は「本物です」ではなく、「これを出したのは確かにこの人で、その人が責任を負います」なのじゃ。

ちゃんとしたメディアが、だまされて偽物を本物と信じて出せば、「そのメディアが確かに公開した」という正直な記録が残る。あとで偽物と分かったとき、逃げられない証拠になるのじゃ。

固定するのは、“真実”じゃなくて“責任”なんだ。

で、結局この写真は「本物」なの？

ここまでやったら、この写真は「本物」だって分かるの？

📌 参考

Content Credentials は来歴データが「真実」かの価値判断を提供しない。

"Content Credentials do not provide value judgments about whether a given set of provenance data is 'true'"

C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

来歴情報だけでは、コンテンツが真実・正確・事実かは分からない。

"…but provenance information alone cannot tell you whether the digital content is true, accurate or factual."

C2PA Explainer 2.4 — 同上

固定するのは、“真実”じゃなくて“責任”なんだ。

そもそも「本物」って、人によって違うんじゃない?

あのさ、そもそも「本物」って、人によって違わない? 面白い“コラ画像”なら、元の写真を探してる人には元が本物で、コラを探してる人にはコラが本物じゃん。

……いいところに辿り着いたな。「本物」はふたつの意味を背負っておる。ひとつは“自分が求めているもの”――これは人によって変わり、共通の正解はない。

もうひとつは“来歴に偽りがないか”――「これは無加工の撮影」「これはそれを加工したコラで、作者はこの人」。この事実は、だれから見ても同じじゃ。Content Credentials が示すのは、こちらだけ。

来歴は、“どっちが本物か”は決めてくれないんだ。

来歴は食べ物の“原材料表示”じゃ。「おいしい」とは書かん。事実を正直に示し、選ぶのはきみ自身。コラが悪いのではない。困るのは「無加工の本物です」と偽るときだけじゃ。

評価は人それぞれ自由。でも、そのもとになる事実は、みんなで守るんだね。

そもそも「本物」って、人によって違うんじゃない?

あのさ、そもそも「本物」って、人によって違わない? 面白い"コラ画像"なら、元の写真を探してる人には元が本物で、コラを探してる人にはコラが本物じゃん。

📌 参考

来歴は、各自が"自分の用途にとって"有用か・信頼できるかを判断するための材料。評価は受け手に委ねられる。

"Content provenance enables them to answer that question, which empowers them to decide how useful or reliable a piece of content is for their use case."

C2PA FAQ — https://c2pa.org/faqs/

補足:「本物の二義(主体で変わる評価/共通の事実)」「来歴=原材料表示」は理解のための整理であり仕様の文言ではない。ただし"評価を受け手に委ねる"方向性は上記 FAQ と整合する。

悪いのではない。困るのは「無加工の本物です」と偽るときだけじゃ。

評価は人それぞれ自由。でも、そのもとになる事実は、みんなで守るんだね。

スクショで消せるなら、意味なくない?

意地悪言うけど——スクショ撮ったら、来歴ごと消えちゃうよね? ぜんぶ意味なくない?

たしかにスクショや再保存で来歴は剥がれることがある。じゃが考えてみよ。剥がして手に入るのは、何じゃと思う?

えっと……来歴のない、ただの写真?

そう。署名も保証も付いてこん。剥がした瞬間に“信用できない側”へ落ちるだけで、元の写真の信頼を奪えるわけではない。“信頼”は盗めんのじゃ。

来歴を剥がしても、“信頼”は手に入らないのか。

さらに最近は、電子透かしや指紋照合を組み合わせて、剥がれた来歴を後から復元する Durable Content Credentials という仕組みも進んでおる。

スクショで消せるなら、意味なくない?

意地悪言うけど——スクショ撮ったら、来歴ごと消えちゃうよね? ぜんぶ意味なくない?

Credentials という仕組みも進んでおる。

📌 参考

Durable Content Credentials は、電子透かしやフィンガープリント（soft binding）を手がかりに、オンライン DB に控えた来歴（manifest）を再発見して復元できるようにする仕組み。透かしとフィンガープリントの併用でさらに堅牢になる。

"If a copy of the manifest data is stored in an online database, you can use a watermark or a fingerprint to find it again." ／ "Combining both watermarks and fingerprints further improves the robustness of the provenance information."

Content Authenticity Initiative — https://opensource.contentauthenticity.org/docs/durable-cr/

補足:「剥がしても信頼は移らない（剥がした複製は無署名＝来歴なしになるだけ）」は、この技術的性質からの帰結の整理。

逆に、画面を撮れば“本物マーク付きの偽情報”が作れない？

👦 じゃあ逆に。画面に映したウソの映像を、本物のカメラでもう一回撮ったら？本物の来歴がついた偽情報ができちゃわない？

👴 ……痛いところを突くのう。そのとおり、これは C2PA が解決できん穴の一つじゃ。正直に言おう。

👴 カメラは「私が、この瞬間、目の前のものを撮った」としか言えん。その“目の前のもの”が、画面に映したウソかどうかまでは、見分けられんのじゃ。

👦 えー！じゃあ意味ないじゃん！

👴 取り違えてはいかん。Content Credentials は最初から「中身が真実か」は約束しておらん。約束は「だれが・いつ・何で」までじゃ。

👴 だからこそ、来歴“だけ”で安心せず、発信者は信頼できるか、話のつじつまは合うか——人の目やファクトチェックと組み合わせて使う。C2PA は万能薬ではなく、その一部なのじゃ。

逆に、画面を撮れば"本物マーク付きの偽情報"が作れない?

じゃあ逆に。画面に映したウソの映像を、本物のカメラでもう一回撮ったら? 本物の来歴がついた偽情報ができちゃわない?

📌 参考

C2PA は誤情報の万能薬ではなく、他の手法と組み合わせて脅威を緩和するもの。

"It is not a cure-all for misinformation, but instead seeks to mitigate against its threats in the digital domain."

C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

補足: 画面の再撮影（アナログホール）で「本物の来歴を持つ誤解を招くコンテンツ」が作れる点は、現行仕様が個別対策として扱っていない正直な限界。

だからこそ、来歴"だけ"で安心せず、発信者は信頼できるか、話のつじつまは合うか——人の目やファクトチェックと組み合わせて使う。C2PA は万能薬ではなく、その一部なのじゃ。

でも、今どこで見られるの?

正直さ……ぼく、来歴マークなんて SNS で見たことないよ。今は関係なくない?

少し前まではそうじゃった。じゃが、いま増え始めておる。たとえば ChatGPT が作る画像には Content Credentials が付くようになり、OpenAI は C2PA 公式の“適合製品リスト”にも載った。

Google も、Gemini アプリに来歴の検証機能を入れて、「今後数ヶ月で Search と Chrome にも載せる」と宣言しておる。ブラウザが来歴を読んでくれる時代が、すぐそこじゃ。

ぼくらが確かめたいときは、どうすればいいの?

だれでも使える検証サイトが、もう公開されておる。画像を放り込めば来歴が読めるのじゃ。EU では機械可読の印を義務づける法律も動き出す。付ける側も、見る側も、そろい始めた。

へえ……! もう始まってるんだ。

でも、今どこで見られるの?

📌 参考

Google は Gemini アプリに C2PA Content Credentials の検証機能を導入し、今後数ヶ月で Search と Chrome にも展開すると表明（C2PA steering committee メンバー）。

"We're also adding verification for C2PA Content Credentials, to easily check if content is an unaltered original from a camera or if it has been modified, and by what tools. This feature is rolling out in the Gemini app starting today, and it will come to Search and Chrome in the coming months."

Google 公式ブログ — https://blog.google/innovation-and-ai/products/identifying-ai-generated-media-online/

OpenAI は C2PA 公式の適合製品リスト（Conforming Products List）に conformant な Generator Product として掲載（画像・動画・音声・PDF、2026-01-28 適合）。公開検証ツール openai.com/verify も提供。

"applicant": "OpenAI OpCo, LLC" ／ "status": "conformant" ／ "productType": "generatorProduct"

C2PA Conforming Products List — github.com/c2pa-org/conformance-public ／ openai.com/verify

---

だれでも使える公式検証サイト: Content Credentials Verify — verify.contentauthenticity.org ／ 規制面でも EU AI Act 第50条(2)（適用 2026-08-02）が AI 生成コンテンツへの機械可読マークを義務化 — artificialintelligenceact.eu/article/50

で、結局……“本物”って、どうなの?

——最初の質問、覚えてる?「この画像、信じていいの?」って。ねえ博士。けっきょくのところ、“本物”って、どうなの?

……それにはな、わしは答えられん。「これは本物です」と決めてくれる機械は、この世のどこにも無い。たぶん、これからも無い。

えー。ここまで来ても、答えは無いんだ。

無い。じゃが——来歴は残る。だれが作り、だれが手を加え、だれが「これは私が出した」と署名したか。その記録は、コンテンツと一緒に、これからもずっと旅を続ける。

答えは出ないけど、“根拠”は、ずっとついてきてくれるんだね。

そうじゃ。信じるかどうかを決めるのは、いつだって、きみ自身。来歴は、そのための根拠を渡し続けてくれる。それで十分なのじゃよ。

で、結局……“本物”って、どうなの？

——最初の質問、覚えてる？「この画像、信じていいの？」って。ねえ博士。けっきょくのところ、“本物”って、どうなの？

📌 参考

Content Credentials は、来歴データが「真実」かどうかの価値判断を提供しない。判断するのは受け手。

"Content Credentials do not provide value judgments about whether a given set of provenance data is 'true'"

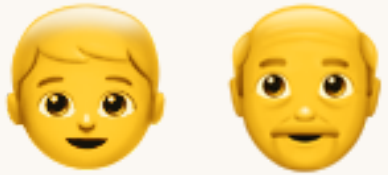
C2PA Explainer 2.4 — https://spec.c2pa.org/specifications/specifications/2.4/explainer/Explainer.html

来歴は、各自が“自分の用途にとって”有用か・信頼できるかを判断できるようにするための材料。

"Content provenance enables them to answer that question, which empowers them to decide how useful or reliable a piece of content is for their use case."

C2PA FAQ — https://c2pa.org/faqs/

そうじゃ。信じるかどうかを決めるのは、いつだって、きみ自身。来歴は、そのための根拠を渡し続けてくれる。それで十分なのじゃよ。



おしまい

来歴は、これからも続いていく。