

# Provenance and Fuzzy Provenance for Text

Marilyn Zhang

## 1. OVERVIEW

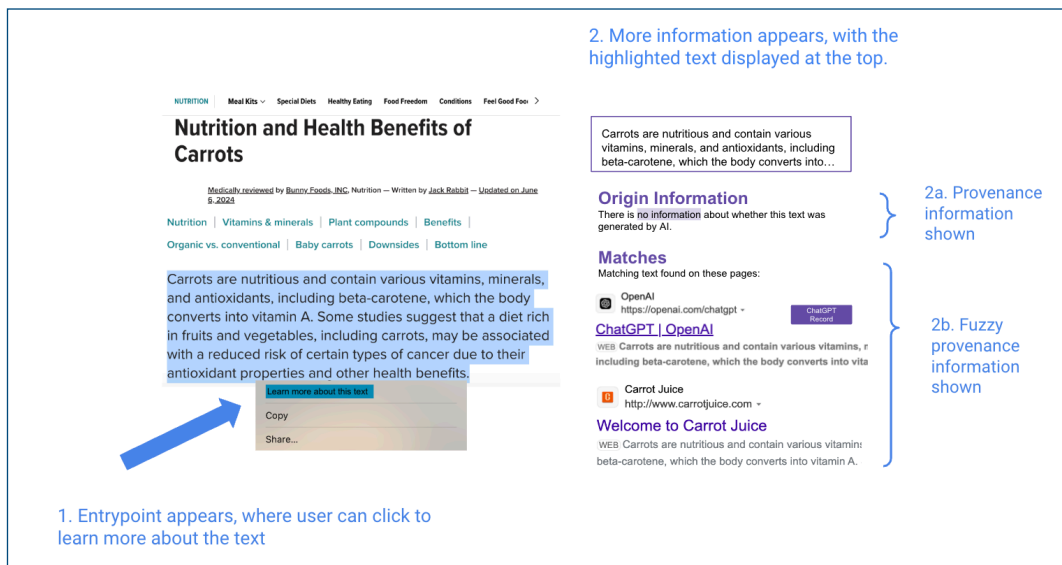
This document operationalizes how digital content hosts can make available (in no more than 1 click) both provenance and fuzzy provenance information about text that users encounter online.

Digital content hosts in-scope for this operational plan would include:

- ▶ Web browsers (e.g., Safari);
- ▶ Social media platforms accessible by apps (e.g., Quora app, WhatsApp); and
- ▶ Websites that are accessible by apps (e.g., New York Times app).

Digital content hosts would be responsible for providing:

- ▶ An **entrypoint** where users can highlight text to learn more about it; and
- ▶ A display, when available, of:
  - ▶ **Provenance** information about the text; and
  - ▶ **Fuzzy provenance** information about the text.



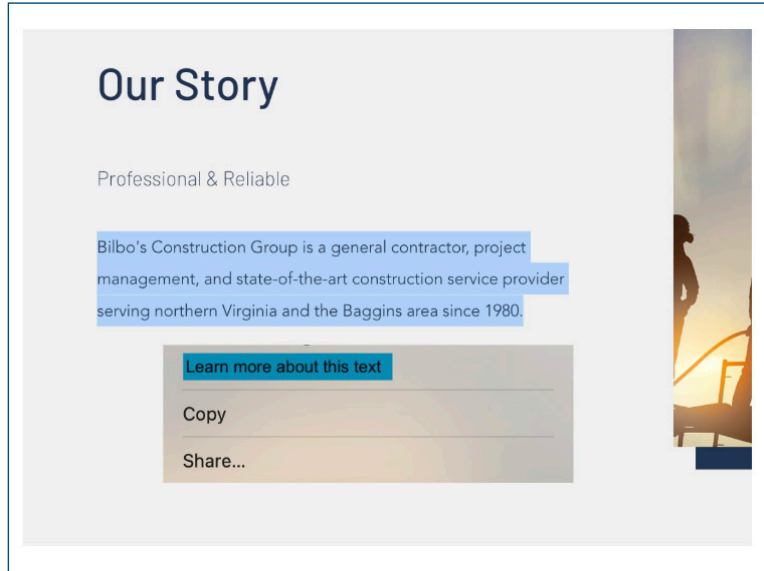
Here is an example overview of what the entrypoint, provenance, and fuzzy provenance information could look like for a piece of text. When a user highlights a piece of text that is sufficiently long, they can click “Learn more about this text” to find more information. First, they see their highlighted text appear at the top of the right hand panel. Then, under the “Origin Information” header, they see information about the text’s provenance. Below that, under the “Matches” header, they see fuzzy provenance information about the text, including exact text matches on the internet and generative AI records, if available.

The rest of this document offers a guide for digital content hosts to provide each of these elements. This guide includes example design layouts based on websites (where web browsers would be the digital content host); similar designs could be adapted for other digital content hosts enumerated above (like mobile applications).

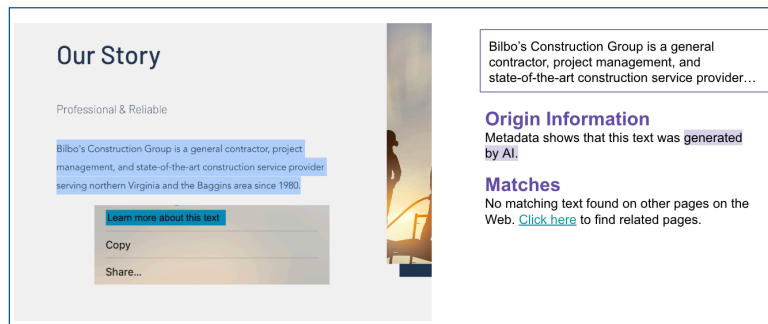
## 2. ENTRYPOINT

An entrypoint should be made available whenever a user highlights a piece of text hosted by a digital content host, and the piece of text contains **at least 15 words**. Specifically, a button saying “Learn more about this text” should appear when these conditions are met, where clicking on the button would lead to a display of provenance and fuzzy provenance information about the text on the same page.

This is what an example entrypoint could look like:



When a user highlights the piece of text starting with “Bilbo’s Construction Group...” the gray hover would appear, which includes the option to “Learn more about this text.” When the user clicks this button, more information about the text would appear in a panel on the right side of the page.



### 3. DISPLAYING PROVENANCE INFORMATION

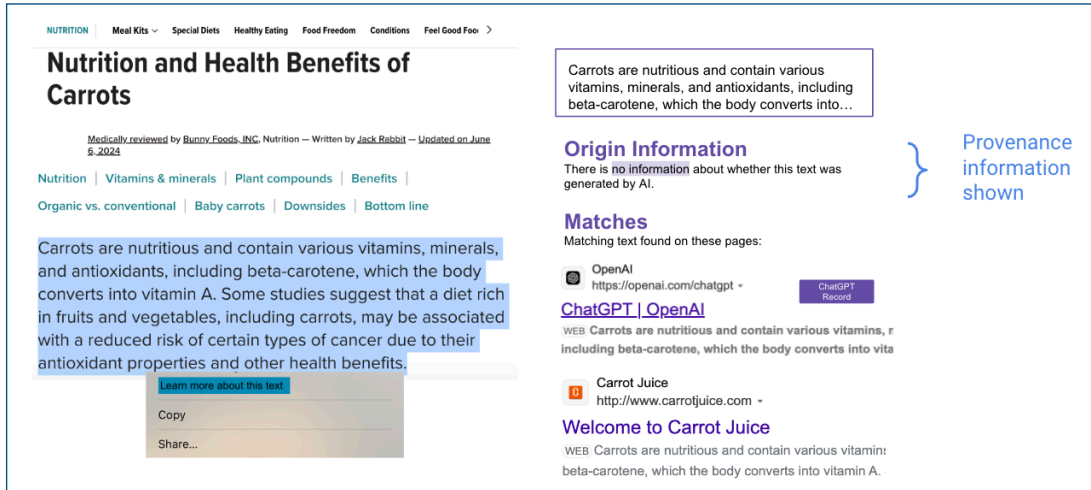
Information about whether the specific piece of text was generated by AI, via watermarking or other information security approaches, and according to [C2PA’s technical specifications](#), should be displayed under an “Origin Information” heading.

Specifically:

Case	Description	Message to display <sup>1</sup>
1	No metadata information, or metadata that is probabilistic predictions with less than 100% confidence intervals, or otherwise inconclusive or contradictory metadata information	“There is no information about whether this text was generated by AI” <sup>2</sup>
2	There is metadata information asserting that the text was generated by AI (including probabilistic predictions that are 100% certain)	“Metadata shows that this text was generated by AI”
3	There is metadata information asserting that the text was not generated by AI (including probabilistic predictions that are 100% certain)	“Metadata shows that this text was not generated by AI”

Examples of cases 1 and 2 are included below.

Case 1, where metadata is not available or inconclusive about whether the text was generated by AI:

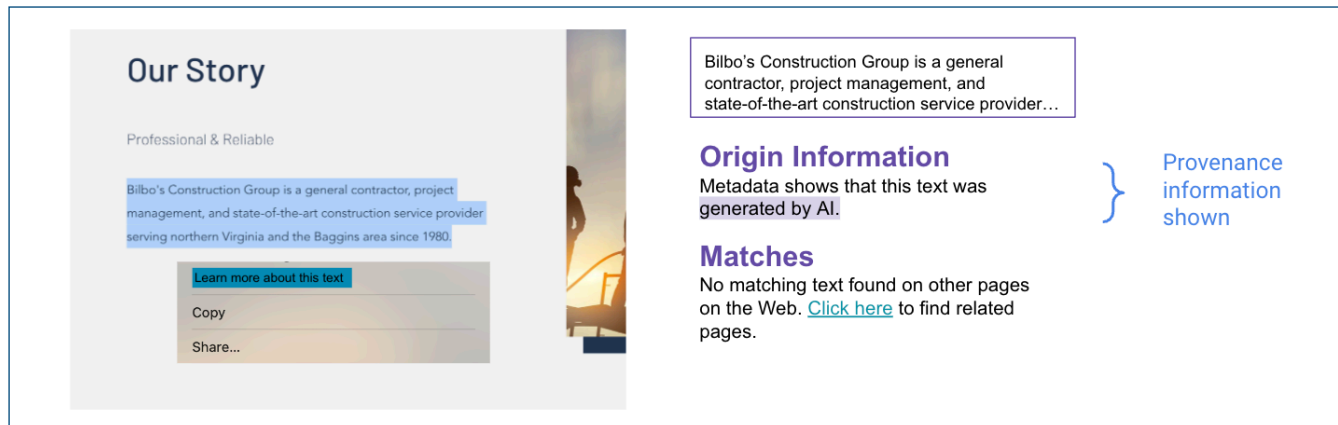


The screenshot shows a webpage titled "Nutrition and Health Benefits of Carrots". The main text, highlighted in blue, reads: "Carrots are nutritious and contain various vitamins, minerals, and antioxidants, including beta-carotene, which the body converts into vitamin A. Some studies suggest that a diet rich in fruits and vegetables, including carrots, may be associated with a reduced risk of certain types of cancer due to their antioxidant properties and other health benefits." Below this text are buttons for "Learn more about this text", "Copy", and "Share...".

On the right side of the page, there is a "Provenance information shown" section. It includes:

- Origin Information:** "There is no information about whether this text was generated by AI."
- Matches:** "Matching text found on these pages:"
  - OpenAI:** <https://openai.com/chatgpt> (with a "ChatGPT Record" button)
  - ChatGPT | OpenAI:** "WEB Carrots are nutritious and contain various vitamins, r including beta-carotene, which the body converts into vita"
  - Carrot Juice:** <http://www.carrotjuice.com>
- Welcome to Carrot Juice:** "WEB Carrots are nutritious and contain various vitamin: beta-carotene, which the body converts into vitamin A."

Case 2, where metadata asserts conclusively that the text was generated by AI:



The screenshot shows a webpage titled "Our Story" for Bilbo's Construction Group. The main text, highlighted in blue, reads: "Bilbo's Construction Group is a general contractor, project management, and state-of-the-art construction service provider serving northern Virginia and the Baggins area since 1980." Below this text are buttons for "Learn more about this text", "Copy", and "Share...".

On the right side of the page, there is a "Provenance information shown" section. It includes:

- Origin Information:** "Metadata shows that this text was generated by AI."
- Matches:** "No matching text found on other pages on the Web. [Click here](#) to find related pages."

#### 4. DISPLAYING FUZZY PROVENANCE INFORMATION

Below the provenance information, fuzzy provenance information should be displayed under a “Matches” heading.

The entries under “Matches” should:

- ▶ Contain only websites that have **exact matches** to the text highlighted by the user: the exact same words in the exact same order, with differences only in punctuation and spelling permissible (to minimize algorithmic curation while maintaining helpfulness);
- ▶ Not include the page the user is already on (to avoid unnecessary duplicative results);
- ▶ Include, for each entry, a preview of the URL, title of the webpage, and snippet ideally including the text surrounding the matched text (to give users helpful context about each source);
- ▶ Be generated automatically (to avoid curation as much as possible), either in-house by the digital content host or by calling a search engine, to include all websites on the internet that contain an exact match;
- ▶ Have generative AI model record results, when accessible on the internet and containing an exact match, clearly labeled (so this is clear to users);
- ▶ Be ordered in some automatic fashion (to avoid curation as much as possible), such as by date (most recent first), with the exception of results consisting of generative AI model records first;
- ▶ Include a link to find related pages (either topically relevant, or matches where a few words are changed) when no exact matches are found on any pages on the internet (to enable users to find paraphrases when exact matches aren't available, since paraphrase matches can provide even more useful context<sup>3</sup>); and
- ▶ Be able to dynamically load more entries if the user continues scrolling past the current viewport (to enable users to find more matches if they want to find more).

Example mocks of 3 cases of fuzzy provenance information are included below:

In the first case, websites with exact text matches are displayed below the subheading “Matching text found on these pages.” None of the results include generative AI model records.

**Quora**  
 Author: Nouman (198 answers, 197.7K answer views)  
 Question: What ginger properties fight against cancer?  
 Answer: Ginger has been considered ten thousand times more Powerful than Chemotherapy for the treatment of cancer.

**Origin Information**  
 There is no information about whether this text was generated by AI.

**Matches**  
 Matching text found on these pages:  
 Vishvas News (https://www.vishvasnews.com/english/health/fact)  
**Fact Check: Ginger is 10,000 times mc**  
 Claim: Ginger is up to 10,000 times more effective than  
 Instagram (https://www.instagram.com/perfectlypurejuices)  
**Perfectly Pure (@perfectlypurejuices)**  
 WEB \* GINGER - Ginger has been considered ten thousand times more powerful than chemotherapy for the treatment of cancer. It is an anti-  
 Facebook (https://www.facebook.com/groups/IndianHealthcareSo)

**Fuzzy provenance information shown**

In the second case, the result includes a generative AI model record. This is shown first and labeled as a “ChatGPT Record.”

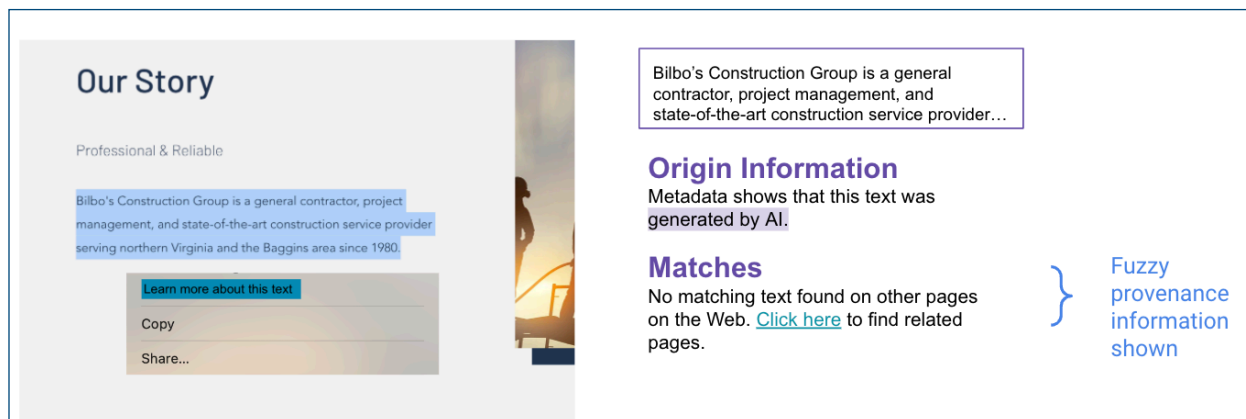
**Nutrition and Health Benefits of Carrots**  
 Medically reviewed by Buniv Foods, INC, Nutrition — Written by Jack Rabbit — Updated on June 6, 2024

**Origin Information**  
 There is no information about whether this text was generated by AI.

**Matches**  
 Matching text found on these pages:  
 OpenAI (https://openai.com/chatgpt) - **ChatGPT Record**  
**ChatGPT | OpenAI**  
 WEB Carrots are nutritious and contain various vitamins, r including beta-carotene, which the body converts into vita  
 Carrot Juice (http://www.carrotjuice.com)  
**Welcome to Carrot Juice**  
 WEB Carrots are nutritious and contain various vitamin: beta-carotene, which the body converts into vitamin A.

**Fuzzy provenance information shown**

In the third case, no pages with matching text are found on the internet. Instead, there is a message that no pages with matching text are found, and an alternative link to related results is provided. When clicked on, the link would lead the user to another page with non-exact-match (but topical) results.



**Our Story**

Professional & Reliable

Bilbo's Construction Group is a general contractor, project management, and state-of-the-art construction service provider serving northern Virginia and the Baggins area since 1980.

[Learn more about this text](#)

Copy

Share...

Bilbo's Construction Group is a general contractor, project management, and state-of-the-art construction service provider...

**Origin Information**  
Metadata shows that this text was generated by AI.

**Matches**  
No matching text found on other pages on the Web. [Click here](#) to find related pages.

} Fuzzy provenance information shown

## ENDNOTES

- 1 The messages displayed are responsive to academic research done on what messages regarding AI generated content are effective and understandable for users. See <https://osf.io/preprints/psyarxiv/v4mfz>.
- 2 Since provenance information for text is often unreliable, it is safer (when not 100% confident) to say "no information" and have the user rely on fuzzy provenance rather than potentially give the wrong provenance information.
- 3 For instance, users or AI may paraphrase from existing text that would show up as a paraphrase match rather than an exact match.

## ABOUT THE POLICY ACADEMY

The Aspen Policy Academy offers innovative training programs to equip leaders across sectors - from tech to climate, science to social impact - with the practical policy skills to craft solutions for society's most pressing challenges.