

## **The Sniffra Project**

Ensuring the future of the internet includes humans.

7/5/2025

## Table of Contents

1- Introduction.....	3
2 - Problem and solution.....	3
2.1 -Problem.....	3
2.2 - Solution.....	4

# 1- Introduction

The contemporary digital landscape over the past 5 years has undergone a profound transformation driven by the rapid popularity of Large Language Models (LLMs) and other Artificial Intelligence (AI) tools. These tools are increasingly capable of generating sophisticated text, images, audio, and video content that poses some difficulty in distinguishing them from human-created works. Despite the potential for increased ease of productivity, it has already raised ethical challenges regarding its transparency, authenticity, and its capacity to create and spread misinformation.<sup>1</sup>

Modern navigation of a world where content on the internet has become interwoven with the outputs of AI systems requires new tools for discernment. Users are increasingly encountering content without clear indicators of its origin, leading to uncertainty and the potential for massive erosion of trust in online discourse and information.

Sniffra emerges as a community-driven solution designed to address this challenge. It is a browser extension that empowers users to collectively identify and verify whether web pages contain AI-generated content. Within this ecosystem, Sniffra aims to provide a real-time "AI-score" for visited web pages, allowing users to form a more complete comprehension of the content they consume. This whitepaper aims to outline the problem and the technologies used in creating a solution in the form of Sniffra.

---

## 2 - Problem and solution

### 2.1 -Problem

The core problem Sniffra aims to tackle is the increasing difficulty for internet users to distinguish between human-authored and AI-generated content. This, coupled alongside the lack of a simple, transparent mechanism for identifying AI's involvement in content creation leads to the unwitting consumption of AI-generated material.

Imagine searching for a brownie recipe, and clicking on one of the first food blogs that populates. Unwittingly, you read the instructions, gather ingredients, and follow the every direction listed on this website. Upon pulling your brownies out of the oven, letting them cool, and having a taste test, you realize that these brownies are equivalent to biting into a sailors hard-tack ration. Dry, bitter, and terrible.

---

<sup>1</sup> <https://www.science.org/content/article/unethical-ai-research-reddit-under-fire>

This is a real world example of how AI content can be harmful. Despite this website's legitimate appearance, it was actually entirely created to abuse search engine optimization(SEO); taking advantage of the systems that search engines like Google or Duckduckgo use to rank and display web results for a user's search. This was entirely to trick you, the searcher for grandmas long-lost brownie recipe, into viewing advertisements for the duration of your baking, on a low quality AI generated website, offering no real value other than its capacity to create really bad brownies and waste your valuable baking time.

Remember, this is just one low-impact example of the proliferation of modern AI content generation. In its current state, there is no telling what these tools are capable of pushing on a wide scale. Imagine, if instead of baked-goods, it was a hot-button political topic with a skewed narrative generated by a nation-state intentionally aiming to misinform. Technologies like these have created a complex set of challenges for the modern internet user and the digital ecosystem at large:

**Opacity and Lack of Disclosure:** Most AI-generated content is published without any clear disclosure of its origin. Users are often unaware they are interacting with content not produced by a human.

**Erosion of Trust and Authenticity:** When users cannot reliably determine the source or authenticity of information, general trust in online content diminishes. This is particularly problematic in areas like news, academic research, and product reviews.

**Misinformation and Manipulation:** AI can be used to rapidly create realistic but false narratives, fake profiles, or biased content at never before seen scale.

**Violations of Art and Intellectual Property:** The rise of AI-generated content raises complex questions about originality, copyright, and the true value of human creativity in the AI age.

**Cognitive Burden:** The average user is already bombarded with vast amounts of information from the internet. The additional cognitive load of constantly questioning the authenticity of every piece of content without adequate tools is unsustainable. This accepts the fact that prevalence of information means less time for critical individual thinking on a single subject matter.

## 2.2 - Solution

In opposition to a single centralized source of information and truth, Sniffra aims to offer a user-centric, community-powered browser extension that enables individuals to:

**“Mark”** the currently visited web page as either "AI-Generated" or "Not AI-Generated".

**“Sniff”** or View an aggregated "AI-score" for the page, derived from the collective input of other Sniffra users.

This approach leverages the wisdom of the crowd to build a dynamic, evolving database of content classifications, providing users with an immediate, accessible indicator of a page's likely origin. Early and accurate contributions will be incentivized, fostering active participation and data integrity.

*White Paper is still a work in progress - Expect updates by 7/12/25*

*Thanks for reading :)*

---