ArtKnights

Requirements Document

26 November 2025

Version 2.3

ArtKnights

Project Sponsor: Andres Sepulveda Morales

Team Faculty Mentor: Bailey Hall

Team Members:
Chandler Silk *(Team Lead)*
Kalyana B R Pabbisetty
Anthony Birk
Ashrith Paranam

Table of Contents

Introduction

Context: Millions of artists have had their personal artwork stolen and used to train AI models without credit or compensation. As artificial intelligence continues to evolve at an unprecedented pace, it becomes increasingly difficult for artists to protect their creative work. Many artists are not equipped with the technical knowledge or tools needed to guard against automated scraping systems that harvest art from the internet. The growing prevalence of this issue threatens not only the livelihood of individual artists but also the integrity of creative industries as a whole. If left unaddressed, it could lead to a future where artists have little control over how their work is used or monetized.

Problem: To push back against this trend, some artists have turned to tools like Nightshade, which attempt to confuse AI training systems. These systems rely on pattern recognition and tagging to learn from data.

Our sponsor, Andres Sepulveda Morales, is deeply invested in defending artists' rights. As the founder of Red Mage Creative Technologies and Fort Collins AI for Everyone, he has been working to educate artists about the risks posed by AI tools and the available defenses. However, his outreach efforts are constrained by scale. As a one-person advocate, his ability to reach artists outside his local community is limited.

Goal: To address this need, we are building ArtGuard, a web-based platform that informs and empowers artists to defend their work from unauthorized use in AI training. ArtGuard will serve as both an educational hub and a reporting tool. Its primary purpose is to expose which websites or services may be harvesting artwork without consent and to provide resources that help artists take action.

The first core feature of ArtGuard will be an educational page filled with curated articles, research, and guides that explain how AI scraping works and what artists can do to safeguard their creations. This will serve as a central resource for learning and awareness.

The second feature is a user-driven reporting system. Artists will be able to submit reports identifying platforms they believe have misused their artwork. These reports will populate a dynamic graph that tracks submission frequency by platform, along with the number of cases that include visual evidence.

We also plan to release a browser extension that displays trust scores and user reviews for websites as they are visited. This feature will let artists assess platform safety in real time, without needing to consult the ArtGuard website directly.

Impact: ArtGuard is designed to support and amplify the work of artists like Andres by scaling education, awareness, and community reporting efforts. It turns a one-person mission into a collaborative platform that helps artists make informed decisions, recognize risks, and take preventive action. By providing accessible tools and reliable information, ArtGuard will reduce the power imbalance between artists and AI developers, fostering a safer and more transparent creative ecosystem.

Problem Statement

Currently Andres conducts educational workshops and discussions about AI and art in his local communities. He shares information via local events, small group sessions, and through social media. This makes it difficult for him to reach a large audience as he is unable to travel to give these talks due to work and time constraints. These artists rely on public awareness campaigns or scattered online sources that become outdated very quickly to learn about AI stealing their art.

Despite Andres' dedication to defending artists' rights and providing knowledge, his impact is restricted by a lack of a scalable infrastructure to educate, connect and empower artists globally. The current process fails to protect artists effectively against the widespread issue of AI art theft.

The main issue with Andres' current ability is the scalability issues. Outreach is limited to Andres' personal capacity and local community events, while also not having a large online presence to spread the message while also allowing those in the community to participate. Furthermore, another prominent issue is the fragmentation of information. Artists have to search across many unreliable or outdated sources to learn about AI scraping while also navigating confusing language that they may not be familiar with. Another pain is the lack of centralized reporting as there is no existing community database that tracks which websites allow art to be stolen and used to train AI. Artists are not able to easily see which website is safe to upload their art. In addition there is a lack of visibility and awareness as many artists remain unaware of tools such as Nightshade or even how AI datasets are collected. Moreover, there are no transparency or accountability tools. Artists have no means to evaluate a website's reputation or risk level in real time nor do they have the ability to see metrics about how safe the websites they use are. Finally, there is also limited community support as artists often act in isolation rather than as part of a coordinated movement as there's no digital infrastructure that connects artists, educators and advocates like Andres.

Overall there are four main problems ArtGuard is looking to solve. The first problem is that artists remain vulnerable to unauthorized AI training and data scraping. Secondly, educational efforts fail to keep pace with the speed of AI development. Third, is the creative community risks losing trust and control over digital art distribution. Finally, Andres' advocacy mission cannot scale or sustain meaningful change without technological support. In conclusion, our missing capabilities are:

- A scalable educational platform for artists
- A community space for discussion, learning, and collaboration
- A centralized, trustworthy database showing which websites are safe or unsafe
- Transparency tools that help artists understand risks in real time
- A digital infrastructure that allows advocacy to grow beyond Andres' personal capacity

Solution Vision

ArtGuard is a web platform that helps digital artists understand risks, report suspected misuse of their work, and collaborate with peers on protection strategies. It combines three pillars into one experience:

- **Learn:** an accessible library of short articles, explainers, and updates about AI scraping and protection techniques.
- **Report**: a submission flow to report suspected misuse (with optional evidence) and to aggregate reports by platform.
- **Discuss**: a modern community forum where artists share experiences, tools, and validate reports.
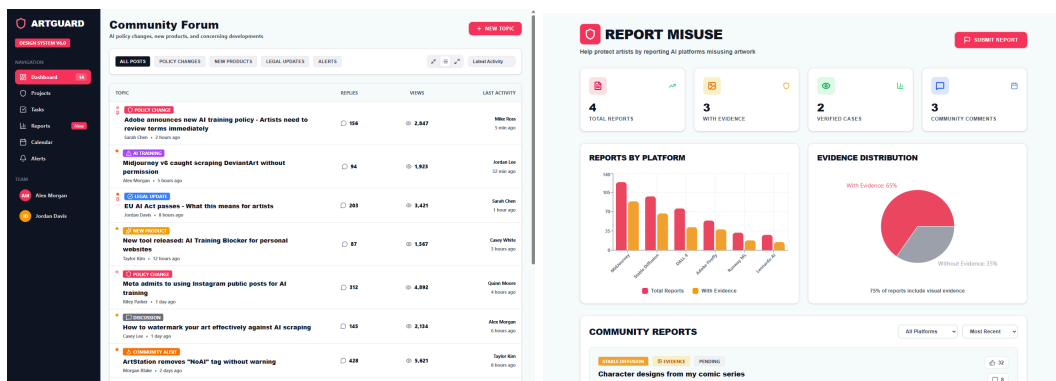

Figure 1: Mockup of Community Forum
Figure 2: Mockup of Reporting Page

Artists often struggle to protect their digital work because there is no proper and structured information or guidance, limited exposure to existing threats, a lack of awareness about protection tools and how to use them, and most importantly, no strong community where artists can connect and speak out together. ArtGuard addresses these issues by providing a dedicated platform where artists can learn, report, and collaborate. Through crowdsourced data, dynamic charts, and an integrated discussion forum, it empowers artists to gain knowledge, and unite as a community to safeguard their creative work.

ArtGuard gathers information from trusted sources like articles, newsletters, user experiences, and a bit of admin research to keep everything accurate and relevant. The system then cleans and processes the data, draws meaningful insights from it, and presents them through clear, easy-to-understand charts and summaries. It highlights potential risks where websites or AI tools might be using artists' work without permission. This makes data collection faster and easier for our sponsor, replacing manual work with clear, real-time insights. Most importantly,

ArtGuard helps even non-tech-savvy artists understand how AI uses their art, how to protect it, what to do if it's misused, and gives them a strong community to share, learn, and speak out together.

Building on the Solution Vision, the next step is to translate ArtGuards high level goals into concrete testable requirements. While the previous section outlined what ArtGuard would like to achieve such as educating artists, enabling reporting, and fostering community, the following requirements section breaks them down into capabilities the product must achieve. Beginning with broad domain level requirements, then transitioning into specific user stories and for our functional requirements a prototype showcasing how these functions will operate.

The following user stories illustrate the requirements in detail, showing how ArtGuards core goals translate into tangible features able to be implemented. Features such as a discussion forum, unified login, reporting dashboards, browser-extension support, trust indicators, notifications, exportable data, mobile responsiveness, accessibility standards and more.

Project Requirements - Functional Requirements

US1: Forum Interaction and Peer Collaboration (Must)

User Story:
As an artist, I want to participate in an online discussion forum where I can share my experiences, learn from others, and collaborate on ways to protect digital artwork from AI misuse.

Prototype:

US1: The Forum is a safe place for artists to share ideas and learn from each other. The prototype shows a discussion created by ArtGuardTeam with three comments by different users, and the ability to leave an additional comment.

US2: Seamless Logins and Web App Transitions (Must)

User story: As a user, I want to be able to transition between the web app and the forums so that I don't have to login each time I go to a different part of the site.

Prototype:

```
https://app.artguard.com
ArtGuard                          Status: Signed out
You are not signed in.
[ Sign in ]
```

```
https://forums.artguard.com
ArtGuard Forums                   Status: Signed out
Sign in to post or reply.
```

```
https://app.artguard.com
ArtGuard                          Status: Signed in
Name        John Doe
Email       johndoe@email.com
  session_id      sess-akzh3mdc
  id_token        jwt-vh2g9epc
[ Go to Forums ]
```

```
https://forums.artguard.com
ArtGuard Forums                   Status: Signed in
Name        John Doe
Email       johndoe@email.com
  session_id      sess-akzh3mdc
  id_token        jwt-vh2g9epc
```

US2: Through the use of Single Sign-On (SSO) authentication tokens, when a user is logged into our web app and clicks on a link to the forums, this token is passed to the forums, and the user is logged in automatically and does not need to sign in themselves. This also works when they move from the forums to the web app.

US3: User Reports of Websites (Must)

User story: As a user, I want to be able to see other peoples reports on ai use on certain websites, so that I can understand at a glance how a website treats art posted on it.

Prototype:

US3: A dashboard that shows a searched website's reviews and allows users to review this site using buttons and updating the average rating.

US4: Extension Support (Should)

User story: As a user, I want an easy and quick way to see how websites handle AI use, so that I can decide whether or not to use that website.

US5: Reputation & Trust Indicators for Users (Should)

User story: As a user, I want users who consistently provide accurate reports to gain a badge so I can know someone is extremely trustworthy.

US6: Web App Notifications (Should)

User story: As a user, I want to be notified when there are replies to comments on forum threads so I do not need to keep checking for new information.

US7: Exportable Report Data (Could)

User story: As a researcher, I would like to be able to export the information into a CSV or PDF so I could use it for presentations, outreach, or academic work.

US8: Community "Issue Tracker" for Platforms (Could)

User story: As a user, I want to see a timeline of reported platform behaviors (e.g., when a site updated its TOS) so I can track changes over time.

US9: AI Detection of Stolen Art (Won't)

User story: As a user, I want to see if artwork has been stolen, and used to train AI.

Reason: Automated detection models are unreliable, ethically sensitive, and too resource-heavy for the project scope.

US10: Legal Advisory or Representation Services (Won't)

User story: As a user, I want legal advice on what to do if my artwork is stolen

Reason: Providing legal advice creates liability concerns; the platform will offer education, not legal consultation.

Project Requirements - Performance Requirements

US11: Mobile Responsiveness

User Story: As a user, I want the mobile version of the web app to load in under 1 second, so that I don't have to wait for it to load.

US12: Reporting Dashboard Load Time

User Story: As a user, I want the dashboard data of reports on the web app to load in under 2 seconds, so that I can quickly review a website's credibility.

US13: Forum Posting Response Time

User Story: As a user, when I create a forum post or reply, the system should confirm submission within 500ms to maintain a responsive, conversational feel.

Project Requirements - Environment Requirements

US14: Accessibility Requirements

User Story: As a user, I want the website to be fully accessible and easy to navigate, so that people of all abilities including those using assistive technologies interact with the platform smoothly.

US14: a fully accessible and inclusive ArtGuard platform designed following WCAG 2.2 AA standards

Potential Risks

Misinformation and Inaccurate Posts – Some individuals will not intentionally convey false information regarding the legal aspects and the AI scraping tools. In the absence of platform moderation, such misguided information will most likely circulate, and the platform will lose credibility.

Inappropriate or Harmful Content – Like any open community, there's a chance of spam, rude comments, or offensive posts. Such behavior will likely discourage participation, from the platform, and make the trusted community feel unsafe.

Data Privacy and Security – Some users will certainly present their work and speak of personal experiences, which can be used to identify them. To ensure the safety of all users, intensive moderation must be coupled with data minimization, clear and easy-to-understand contracts, and privacy policies.

Low User Engagement – Inactive participation of users, especially in rating, discussion, and wtools forums, will directly impact the system's overall purpose, main impact, and effectiveness.

Scalability and Maintenance – As the community grows, managing users and data will become more complex. Regular reviews will be needed to keep everything running smoothly and efficiently.

Project Plan

Taking into account the user stories we have collected, the problems and risks we've identified, and the solution we've envisioned, it becomes easier to figure out the timeline that the project will operate on. We've identified several core issues that our project aims to solve, so the timeline is built around those solutions. The Gantt chart shown in Figure 1 below shows our initial plan for project milestones and workflow.

As stated before, the three main uses of our web app are creating and displaying articles, compiling and displaying user ratings, and a forum environment for user discussions. Our milestones center around these uses. The first milestone will be the creation of a simple prototype, which will also act as the initial wireframe for the web app. This prototype does not need to be fully functional, but it needs to show an early idea of what the web app will look like.

From there, we begin working on the main use cases. First, a workflow for the articles will be established using the Sanity CMS, allowing for the creation, publishing, editing, and deletion of articles by users. Once that is finished, we move onto building the full web app skeleton using Next.js. This allows us to render the website and display its contents. After that, we will set up our Postgres database in order to store the ratings information, which we will build the API off of. Discourse integration will allow us to host a user forum, and will allow users to switch between the forums and the web app freely. Once we have these parts up and running, we can build the charts using Charts.js, and integrate the ratings data into the different parts of our web app.

This gives us a firm foundation to begin testing the web app. During the entire testing process, we will be performing tests to ensure responsiveness and security throughout our website. Initially, this will be with custom datasets. Once the website visuals are fully developed, we can move on to usability testing for users, which allows us to ensure that our user's needs are met. Once this is reached, we will prep the web app and stage it for release. Once released, we can see how it handles active use, and deal with any issues that arise.

ID	Task Name	2025			2026			2026	
		10	11	12	01	02	03	04	05
1	Web App Wireframe/Prototype Creation		▓	▓					
2	Establish Sanity CMS Workflow				▓				
3	Create Next.js Web App Skeleton				▓				
4	Ratings Model & API				▓				
5	Discourse Forum Setup					▓			
6	Ratings & Chart Integration					▓			
7	Performance & Security Testing						▓	▓	
8	Implement Web App Branding						▓		
9	Usability Testing for Active Users							▓	
10	Prerelease Polish							▓	
11	Web App Launch Phase								▓

Powered by: onlinegantt.com

Figure 1: Project Gantt Chart, Tasks are shown in which months it will be worked on

Conclusion

This document outlines the complete vision of the ArtGuard project and how it will be designed, built, and maintained. It starts with the Introduction, escalating the problem of unreasonable use of creators' work by AI systems and explaining the need for protective systems like ArtGuard. The Problem Statement pinpoints the issues of unawareness, cluttered and fragmented information, and scarce community support that burden the artists. The Solution Vision depicts ArtGuard as a web platform where artists can educate themselves about the risks of AI, report the infringement of their art, and engage with others in community discussion forums.

The Project Requirements section articulates the expected core functionalities of the system such as forums, community reports, extensions, and support, and outlines the expected operational and environmental conditions. The Potential Risks section lists the challenges of misinformation, toxic content, privacy, disengagement, scalability issues, and potential mitigation strategies. The Project Plan details the phased approach of developing, then testing the system to meet a defined set of usability and access criteria for the secure, responsive platform before launch.

ArtGuard demonstrates the potential for a structured approach to building a secure digital environment for creatives. It outlines the value of defensive measures regarding protected work, delivers a practical technical approach to accomplish this, and illustrates the ways ArtGuard assists artists in education, reporting, and the digital preservation of collaborative art.

RED MAGE

TO: **Project Team / Stakeholders**
         NAU Capstone Faculty

FROM: Andres Sepulveda Morales
           Founder, Creative Technologist

DATE: December 1, 2025

RE: **APPROVAL OF ART GUARD REQUIREMENTS DOCUMENT**

This memo serves as formal confirmation of my review and approval of the Requirements Document, titled ArtKnights Requirements Document Version 2.3, dated 26 November 2025.

As the Project Sponsor, I confirm that the documented requirements align with the strategic objectives and business needs for the Art Guard project. I authorize the Project Team to proceed into the next phase (e.g., Design, Development) based on the specifications outlined in this document.

Please do not hesitate to contact me at andres@redmage.cc or via phone at (970)286-3041 should you have any questions or concerns.

Cheers,

Andres Sepulveda Morales
Founder, Creative Technologist