

Ansanm: AI Art Assistant

Design Document

Ansanm (Haitian creole for “together”, from the French “ensemble”) is [ethically designed](#), open-source software that increases accessibility of visual storytelling and provides simple, thoughtful interaction with an artistic assistant AI.

The Ansanm AI will help artists quickly construct an initial concept based on a story, text or poem which is then available for further refinement using smart search and comparison tools. The goal is to make the creation of visual media accessible to people of all skill levels and allow for diverse storytelling. For example, given a story, a visual essay with correct timing and smooth transitions will be constructed automatically from the artist’s image library, internet content and AI generated content. Images can then be replaced, removed or inserted as needed. Use cases include:

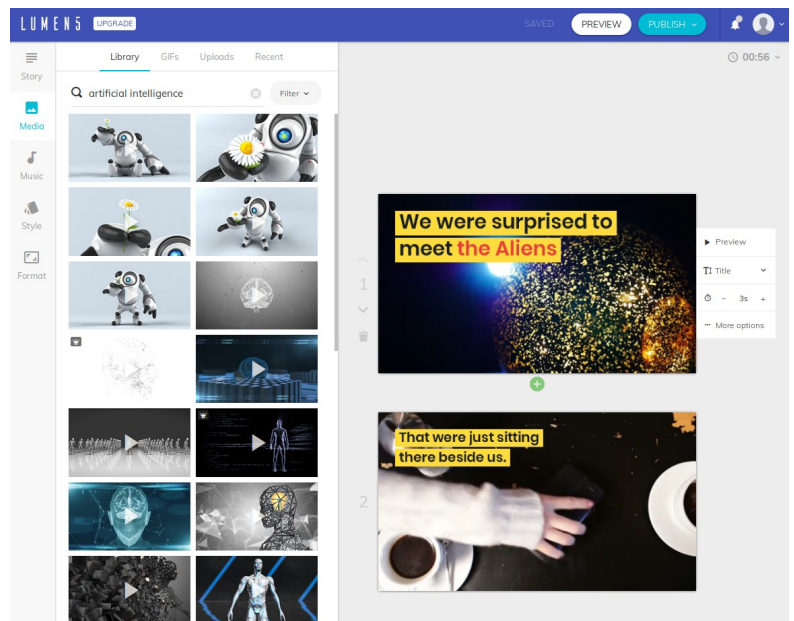
- Music videos, video essays and video installations
- Photo slideshows with background music
- Poetry with integrated visuals and other talks and presentations

Ansanm’s AI assistant can translate one medium to another, the simplest being text to image. Artists can then explore new mediums, opening up avenues that were not possible before because of a lack of training, interest, or physical ability.

Ansanm will help artists explore a new form of art making that centers around human-machine collaboration and curation. The AI assistant will find or create many options for the artist-as-curator to choose from, who can then guide the AI to modify that selection until it finally fits their artistic vision.

Basic interface components:

- Content search (images, video, music):
 - Content comparison and filtering
 - Related-search suggestions
- Text content:
 - Import
 - Editing
 - Placing into sequence and layout
- Sequence and transition editing:
 - Sequence content selection
 - Sequence order
 - Length adjustment
- Load / import AI model
- Music and audio selection and visualization



There are some existing for-profit businesses using some basic AI for presentations and video essays but they are strongly biased towards business and monetization on social media platforms. None are open source, most are not freely available, and all have unknown privacy concerns. They do provide an example that the technology exists and illustrate the basic features and user experience:

<https://lumen5.com/>

<https://www.rawshorts.com/>

<https://spark.adobe.com/>

<https://www.animaker.com/>

User Interface Example

Example of Lumen5 user interface, of a poem whose text has been analysed, keywords found, and images suggested based on those keywords. Text from the poem was then added to the images.

The screenshot displays the Lumen5 user interface for creating a video. The top navigation bar includes the Lumen5 logo, an 'UPGRADE' button, and 'SAVE', 'PREVIEW', and 'PUBLISH' buttons. A left sidebar contains menu items for 'Story', 'Media', 'Music', 'Style', and 'Format'. The main workspace is divided into a text editor on the left and a video preview area on the right. The text editor shows a poem with lines highlighted in blue. The video preview area shows two video frames: the first frame has the text 'We were surprised to meet the Aliens' overlaid on a starry space background, and the second frame has the text 'That were just sitting there beside us.' overlaid on a coffee shop scene. A timeline at the bottom indicates a 00:56 duration.

Use Cases

- **Amy is a poet** and wants to disseminate her work on YouTube and StoryFire:
 - She has no video editing skills or money for expensive video editing software
 - Amy uploads the text of her poem:
 - The text is parsed for key words which are in turn used to start a smart image search
 - For each keyword:
 - Amy is presented with a series of images and videos
 - She marks some that definitely do not work and others that she is interested in
 - The smart search refines her options
 - Amy selects an image or video that fits that part of the poem
 - Amy locks together which word in her poem is associated with that image
 - Amy uploads a voice recording of herself reading the poem
 - Ansanm matches timing of the images to the timing of the voice
 - Amy nudges the timing of the images, tweaking it until she is happy
 - Amy exports a video that she uploads to the streaming services
- **Bob needs a quick storyboard for a short film:**
 - Bob quickly types out some ideas and brainstorms some themes and uploads that into Ansanm
 - As above for Amy, using the smart search feature he selects images that match
 - Bob needs to add a lighthouse to a beachfront image
 - He selects the area he wants the lighthouse and selects an example lighthouse from other image, Ansanm **creates a set of example images that combine both images**
 - Bob selects his preferred image
 - Bob has an idea for the main character's look
 - He searches through a set of AI generated faces to find one that is close
 - He uses that as an example for the AI to generate more similar faces and chooses one of those
 - Bob has Ansanm replace the face of one of the people in the storyboard with the face of his character
 - Bob exports a video animatic of his story board
- Carlos wants his Ansanm video **slide show to sync up to his choirs' singing** in a live concert:
 - Carlos uploads a recording of his choir singing the piece
 - Ansanm does speech recognition and displays text in time to the audio
 - Carlos corrects any errors
 - Carlos locks together the text and audio pairings with images
 - When the choir sings live, Ansanm is fed the live audio and **displays the visuals in time to the music**
- Dina wants to **tell her story about the pandemic:**
 - The story loops on itself, each day feels like "Groundhog day" repeating itself but presenting subtle new choices
 - Dina uses her phone to take images and video of her apartment and her things
 - Dina **creates story branches of sequences of images and videos** and marks items in the images that when clicked on take the user to a different branch
 - Dina changes the mode to "Advanced" which lets her edit the project with the entire capabilities of Godot and writes some code that tracks how many clicks the user makes and after 20 clicks goes back to the root "sleeping" story branch and resets the click count
 - Dina adds some code so that after 10 "days" in the story a new branch starts containing the ending
 - Dina exports her project as a phone app, desktop software and a web page

Feature plan

In Phase 1, all features are required for a minimum viable project. In Phase 2 features can be added optionally as funding allows.

Phase 1 use cases:

- Photo slideshows with background music
- Poetry with integrated visuals and other talks and presentations
 - E.g. allowing for art similar to the **Creo Animam** concert
- Music videos, video essays and video installations

Phase 1 features:

- Import and display of text, images, videos and music
- Timeline and editing of content sequences
- Assisted image, video and music search:
 - Search multiple copyright-friendly (creative commons, etc) databases at once
 - Assisted rights and credits, i.e. automatic credits generation
 - Content processed by AI for further tagging and categorization
 - Automated content searches given text or lyrics as input
- Automated sequence/timeline visual content generation from text content
- Export to video

Phase 2 use cases:

- Video/slideshow synchronized accompaniment (to speech and/or music)
 - E.g. improving on the art similar to the **Creo Animam** concert
- Live interactive VJing
 - E.g. allowing for art similar to the **Coding Chaos** concert
- Interactive visual novels and storytelling games

Phase 2 feature options:

- Improved smart searches
 - Marking images as good/poor fits and re-adjusting search
 - Searching for similar images
 - De-duplication of content (e.g. hide alternate versions or copies of images)
- Interactivity (support for keyboard, mouse, game controllers, webcam input, etc)
- Branching sequencing (allows of different flow based on input, etc)
- Timing controls through audio input (i.e. advance to next image/video based on audio input)
- Dialogue systems for visual novels
 - Import of dialogue from narrative systems like Twine, Yarn, etc
- Collaboration and sharing
 - Easy publishing of your source code and images and access to others
 - Sharing of AI models
- Image, video and music generation (see examples below)

Dissemination: Visual media made by myself using Ansanm will be made at the end of Phase 1 and 2. This will be used to promote the tool online as well as in in-personal local demonstrations to local Toronto arts and gaming communities; including Hand-Eye Society, OCAD, Godot Toronto user group, Site 3, Hacklab, etc.

Detailed work plan

(Each bold heading represents 1 month of work.)

PHASE 1 (8 months)

Building infrastructure/plugin-ins for basic functionality

- Research Godot plugins
- Set up development environment
- Test godot plug-in system
- Basic UI for image selection (plugin)
- API image search prototype
- Basic UI for image/slide sequences (plugin)
- Local asset library
- Import local assets
- Image tagging
- Search by tag
- Basic UI for script / text (plugin)
- Text import
- Connecting text to image/slide sequences

First minimally functional prototype

- Slide / sequencing engine
- Create project website and documentation
- Playback of sequence
- Simple transitions
- Testing export to Windows, Mac, Android, iOS
- Test sequence / demo
- Text on slides
- Text positioning / layouts

Video import / export support, basic music support

- Video import: mp4 to ogv auto-import
- Export to video
- Video editing: slice/trim video
- Music import
- Music selection UI
- Music playback
- Volume adjustments / fades / etc

Slideshow controls and other basic input, general UI improvements

- Input support
- Simple slideshow controls
- Improved slide transitions
- Transition selection UI
- Improvements to timeline / sequence UI
- More search engines
- Improvements to search UI (filtering, comparison)
- Improvements to music/sound support

AI research and integration of machine learning and Godot

- Research integration
- Integration of godot and ML
- Run models inside godot (editor)

Basic AI assistance: smart search

- Research best ML models to use
- Automated search based on text
- Automated scene length (based on text)
- Image / video similarity (de-duplication)
- UI improvements for AI integration

AI development: understanding content and automated content selection

- Object recognition in images
- Object recognition in videos
- Automatic tagging
- Detect qualities of images / videos
- Automated sequence generation from text
- UI development and improvements

Creating with tool, release v1

- Design and scripting
- Image, video, music searching and selecting
- Editing / sequencing
- New code / features needed for completion
- Code / software release to public
- Demonstrations to local arts and gaming communities

PHASE 2 (up to 10 months, in approximate order)

Refinements based on user testing

- Improved usability based on feedback from users

Collaboration and multi-users

- Simplify multiple users working on one project
- Packaging project code and assets for ease of sharing

Real-time usage, similar to *Creo Animam*

- Listen to audio input to control timing of visuals

Non-real-time AI content generation

- Create images and videos from text or sketches
- See below for examples

Content sharing and community building

- Easy sharing/finding of projects, assets and AI models
- Website and community forums

Interactivity and non-linear flow

- Sequences with branches and loops
- Import / creation and basic support of dialogue systems

Story assistant

- Help generate text, and suggest improvements
- Tutorials and questions about story content and structure

Performance optimizations, shaders and effects

- Shaders allow for very fast GPU based visuals/effects

Real-time usage, similar to *Coding Chaos*

- Real-time AI content generation in response to input

Polish, UI improvements, bug fixes, v2 release

Advanced AI assistant feature examples

Examples below of existing AI that can be integrated into Godot:

Image generation from text:

Example from [AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks](#) paper:

this bird is red with white and has a very short beak

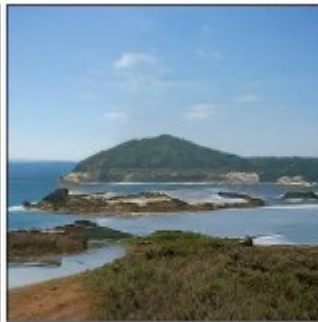


Example images generated from given text from [BigGAN by DeepMind](#):

“Dog”:



“Island”:



“Butterfly”:



“Hamburger”:



Real-time image generation from webcam: Example from my work on the *Coding Chaos* concert: (image on left is view through webcam, right is generated by AI in real-time)

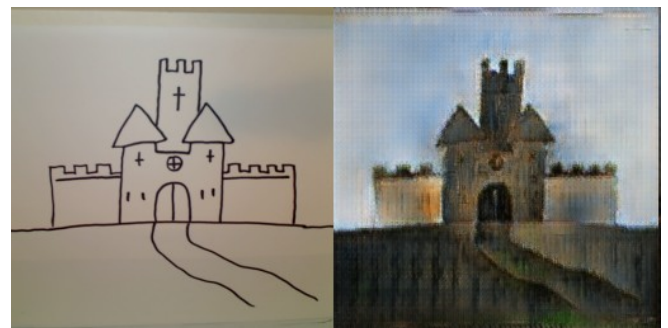
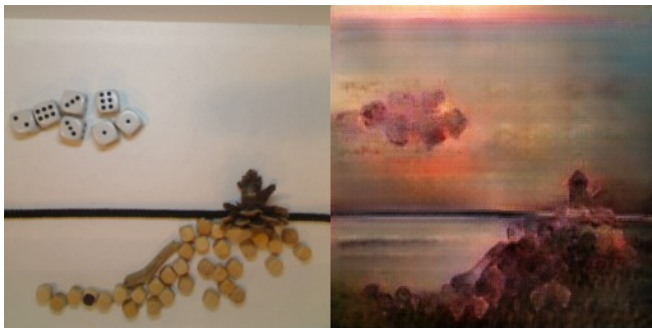
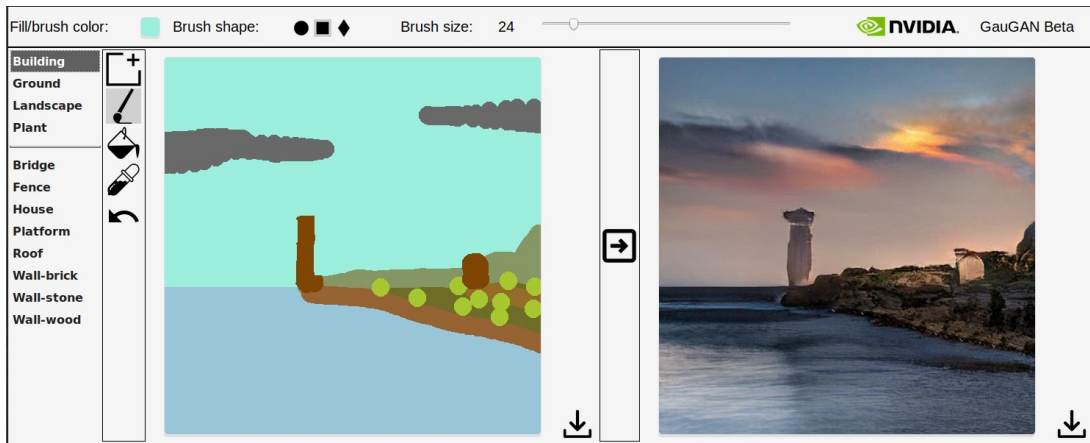


Image creation from sketch: Example from <http://nvidia-research-mingyuliu.com/gaugan/>



Example from [SPADE by NVIDIA](#):

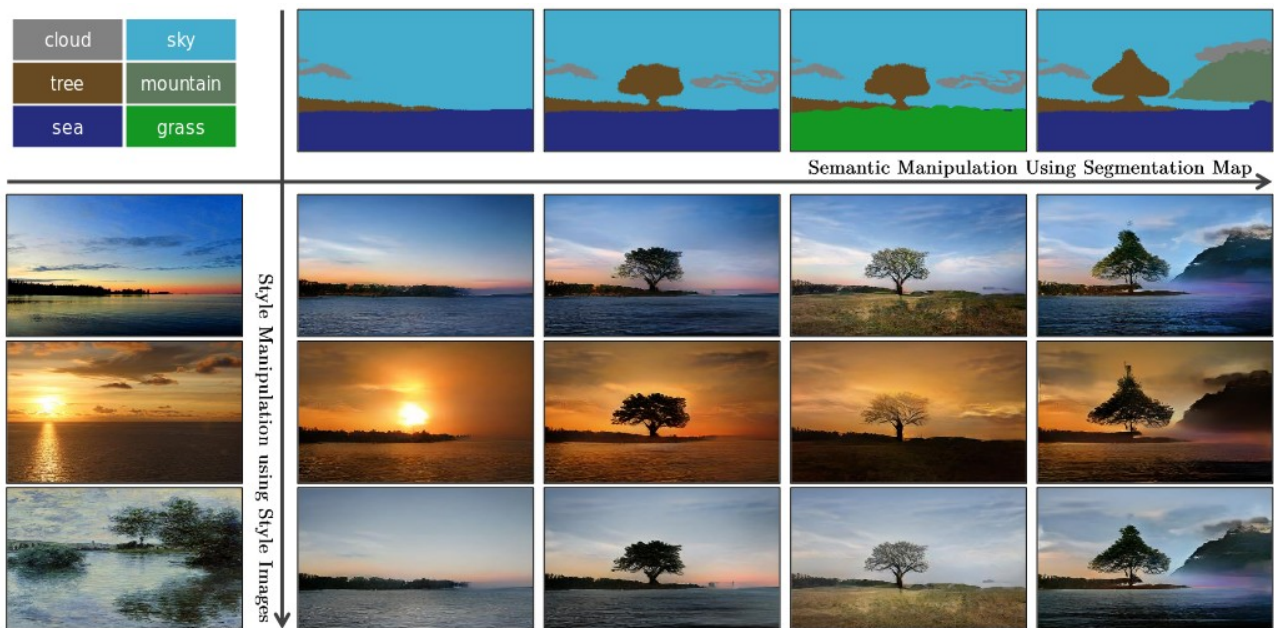


Figure 1: Our model allows user control over both semantic and style as synthesizing an image. The semantic (e.g., existence of a tree) is controlled via a label map (visualized in the top row), while the style is controlled via the reference style image

Technical details

Ansanm will be built on existing open-source software (Godot game engine) that I've previously used to create game prototypes. Its strengths include compatibility with multiple operating systems, extensibility, and ease of use for novices and depth for experts. I will build a plug-in to integrate existing open-source ML algorithms and connect to RunwayML (an open platform for ML sharing and experimentation). This reduces the scope of the project considerably, building on existing tools and allowing Ansanm to focus on accessibility and art making.

All of the code will be open source.

Collaborators

I have included Chris Tihor as an expert consultant on narrative design, as I have less expertise in this area. Chris is a professional game writer and designer, with experience writing for a variety of game projects from indie to AAA.

Part of the inspiration for this project comes from direct experience working with artists whose current tools did not suit their creative practice. Additionally, I purposefully set out to find groups and partners with experiences outside my own but had aligned purposes or goals - particularly around accessibility of art, art-making, and/or machine learning and an emphasis on (visual) storytelling. I have invited these artists and arts groups to participate in design, testing, and use of Ansanm, a process that will continue throughout the project. Ansanm addresses the difficulty in creating digital art, directly lowering barriers and working towards an accessible, inclusive community of digital art making. By incorporating feedback from a diverse set of creators I hope to further break down barriers.

Collaborators participating in design discussion and beta testing are doing so on a volunteer basis, as typical in open source community software projects, with the expectation that they have free access to the Ansanm software. I will also provide small honorariums for each.

Distribution and Dissemination

Ansanm will be distributed online, on my own personal website as well code hosting services. Videos created with Ansanm will be hosted on video streaming services and be submitted to local festivals (Luminato, Nuit Blanche, Toronto Storytelling Festival).

I will be engaging with and promoting to local Toronto-based groups:

- The Hand-Eye Society (indie gaming)
- Interaccess, Music Gallery and other new media arts groups
- local / online Godot (and software) user groups
- arts institutions: OCAD and MOCA
- community arts orgs: Artscape and the ArtsHeart Community Art Centre

As well as working with supporting artists to get their work disseminated in their local communities.

One of my strongest motivations behind the open nature of the project is that this encourages the development of potential future collaborators: either software developers who extend the software itself or other artists I can collaborate with on content produced with the software. Ansanm is a testbed for promoting digital artistic practice and collaboration and provides greater insight to what techniques (both in software and community building) are most effective.

Why Ansanm?

Despite ML research growing rapidly, artistic exploration of ML is constrained by technical requirements. Projects like RunwayML (which will be integrated) are starting to address this problem, but Ansanm focuses on visual storytelling to further increase accessibility.

Visual storytelling is a critical tool for mass movements for social change: it is often more impactful and accessible than text. I hope this project gets used by people who otherwise might not have the skills necessary to tell personal, compelling stories that help connect people and drive positive social change.

Once familiar with the simpler visual storytelling aspect of the software these users can then start experimenting with the more advanced features, encouraging even more people to become visual artists and experiment with machine learning.

I hope to challenge the cultural norms around digital art making and encourage ethical, sustainable creator-centered software and practices. This project hopes to introduce these concepts to Ansanm's users and in turn encourage their audiences to become artists themselves and create their own art with Ansanm, exposing them to the same concepts.

Promoting the idea that great art builds on what has come before, Ansanm will integrate decentralized peer-2-peer technology to help users discover other Ansanm-created art. This makes attribution essential. Ansanm helps automatically track attribution and includes links to projects and contributors, creating connections between them.

Artist's Statement

In 2001, my approach to art changed due to personal and world events. I began full-time art-making, collaborating with Yoko Takashima, and researching history, power, economics and the free software movement. I decided I could make the most impact creating art promoting authorship, equity, privacy, free access and collaboration, and advocating for the abolition of digital property and advertising.

Since my 2015 *Creo Animam* concert, with Vox Choirs Toronto, I increasingly focused on the dangers and opportunities around artificial intelligence and automation. In 2019, I helped Spectrum Music composers integrate machine learning to generate music, lyrics and created my own ML-generated visuals for the *Coding Chaos* concert. Most recently I have worked with Eve Egoyan and Jason Baerg, helping them integrate ML into their latest work.

This inspired me to find new opportunities for artists and audiences to deeply interact with ML, specifically in an art-making context. The more I researched AI and its use, the clearer it became that humanity's disregard for the consequences of unrestricted profit and power had created two existential threats: climate change and the development of harmful AI. In both cases, artists can work to deepen the public's connection to these issues, increasing their capacity to make responsible decisions.

Like Joseph Beuys ("everyone is an artist"), I believe in universal human creativity and art-making to bring about revolutionary change. Over the next few decades as we transition to a fully automated economy, life-long education and relationship building must replace labour in providing meaning to human existence.

This is a once-in-the-history-of-a species challenge requiring an understanding of machine thinking that is

democratic, personal and skeptical of corporate and state narratives. Ansanm presents AI as collaborator and muse: helping you to express yourself authentically. The stories created will teach us about our disparate lives in difficult times, and eventually teach our machines. Human knowledge is often shaped as a narrative, so ultimately, an AI that seeks to know us must learn human storytelling so it can communicate. Tools like Ansanm are the first steps toward first contact.

Ansanm is the embodiment of my hope that accessible, ethical, free software for storytelling and art-making, enriched by the latest research in machine learning, can help us communicate and make change together.

Appendix: Algorithmic Bias in Art

As “software eats the world” (Marc Andreessen in 2011), especially a world driven by corporate profits and network effects, the ethics of the role algorithms play in our lives needs to be questioned rigorously and unpretentiously¹. Algorithmic bias refers to systemic and unfair discrimination in a computer system. Until recently, a widely held myth claimed that software, particularly AI trained on large datasets, was not biased.

All software is biased; writing software is the act of recording the decisions of the biased humans and often corporate entities responsible for its creation. This includes both machine learning algorithms and image and video editors typically used by artists. In addition, the training data used in machine learning is curated by humans, further adding bias.

The impact of algorithmic bias changes in an art context. Unlike in commercial or government contexts, engagement with art is voluntary and art is understood to be subjective. Thus biased, i.e. subjective, results can be appreciated as such and even used to critique bias.

Like a human collaborator, AI can only produce art based on what it has experienced, and the data used to train the machine often represents an infinitesimally narrow slice of human experience that is nonetheless too vast to experience in a single lifetime. For example, DeepMind’s Go and chess playing AI, AlphaZero, was trained by playing millions of matches against itself. It is considered the best Go, shogi and chess player in the world amongst all humans and machines and is inspiring changes to how humans play. Machine subjectivity, potentially as unknowable and inexplicable as the motivations of any artistic collaborator, can simultaneously inform us from a position of superhuman mastery and utter alien ignorance.

Embracing and acknowledging bias as subjectivity, through transparency of the process and the software, leads to a shared, vulnerable exploration. An open source approach allows for ethical software-based art, where bias is assumed, but is not inescapable. AI that can be inspected, modified or replaced is malleable; indeed, far more so than the bias of the creators themselves. Openness of code, data and process leads to learning and opportunities.

One important note, as yet a future consideration; there may be no ethical construction of a sentient machine intelligence – not a Terminator threat, but a Frankenstein-ian failure. It puts the creator in the impossible position of being a parent to an alien child that may pose a danger to itself or others; a parent required to decide on the captivity and/or study of this child of unknown capabilities and needs.