Reality-Aligned Intelligence (RAI) for Creativity

Ethical AI-Assisted Creativity, Origin Honesty & the Blue / Green / Yellow Labels

Author

Niels Bellens
Independent researcher – AI, education & neurodiversity
Antwerp, Belgium

Version

Draft v0.1 – December 2025

Abstract

This whitepaper proposes a Reality-Aligned Intelligence (RAI) framework for ethical AI-assisted creativity. As generative models become embedded in writing, art, music and design, markets and audiences increasingly struggle to answer a simple question: *who actually created this?* Current debates are polarized between "all AI art is theft" and "AI is just another tool", while platforms often obscure origin and process.

The RAI Creativity Framework offers a middle path based on **ontological honesty**: telling the truth about what a work is, how it was made, and whose creative agency it expresses. It introduces (1) a **Creativity Hierarchy** (creation → co-creation → generation → replication), (2) three visible **origin labels** — 🔵 *Blue Seal* (human origin), 🟢 *Green Seal* (human–AI collaboration), 🟡 *Yellow Seal* (AI-generated), and (3) domain-specific protocols for visual art, text, music, film, games and cultural heritage.

Combined with the broader RAI governance and auditing stack, this paper sketches how origin honesty can be measured, signaled, verified and enforced in real creative ecosystems — in ways that protect artists' rights and dignity, support fair markets, and still allow responsible, transparent use of AI tools.

Intended audience
- Creators (writers, artists, musicians, designers, game makers)
- Platforms, publishers, labels and galleries
- Regulators, collecting societies and standards bodies
- AI tool builders and product teams
- Researchers in AI ethics, IP, media and cultural policy

Table of Contents

4. RAI Creative Origin Labels (Blue / Green / Yellow)

4.1 Design goals for origin labels

4.2 🔵 Blue Seal – Human Origin

4.2.1 Criteria

4.2.2 Typical use cases

4.2.3 Rights and expectations

4.3 🟢 Green Seal – Human–AI Collaboration

4.3.1 Criteria

4.3.2 Process transparency and logs

4.3.3 Example royalty splits and crediting

4.4 🟡 Yellow Seal – AI-Generated

4.4.1 Criteria

4.4.2 Platform duties (clear labeling, no fake "human" claims)

4.4.3 Acceptable uses and limits

4.5 How labels appear in interfaces, catalogs and contracts

4.6 Relationship to the wider RAI metrics and auditing stack

5. Five Principles of Creative Integrity

5.1 Transparency: tell the truth about origin and process

5.2 Consent: respect style, data and community boundaries

5.3 Attribution: credit humans and sources properly

5.4 Human Presence: preserve real human agency and meaning

5.5 Creator Dignity: protect livelihoods, reputation and future work

6. The Creativity Hierarchy in Practice

6.1 Purely human creation (no AI tools)

6.2 Human creation with non-generative tools (DAWs, Photoshop, etc.)

6.3 Human–AI co-creation (assists, drafts, variations, brainstorming)

6.4 AI-primary generation with human curation

6.5 Replication, style mimicry and deepfake creativity

6.6 Mapping hierarchy levels to Blue / Green / Yellow labels

7. Domain Protocols: Overview
7.1 Why different domains need different rules
7.2 Shared constraints across domains
7.3 How domain modules link back to RAI auditing metrics

8. Visual Art & Illustration
8.1 Style consent, exclusion lists and licensing
8.2 Training data transparency for visual models
8.3 Labeling obligations for galleries, platforms and buyers
8.4 Example scenarios and edge cases

9. Text, Literature & Knowledge Work
9.1 Ghostwriting, AI drafting and human editing
9.2 Academic integrity and student work
9.3 Non-fiction vs fiction vs marketing content
9.4 Blue/Green/Yellow in publishing workflows

10. Music, Voice & Audio
10.1 Voice cloning and synthetic performances
10.2 AI-assisted composition, mixing and mastering
10.3 Royalties and splits for AI-assisted tracks
10.4 Posthumous albums and estate rights
10.5 Emotional manipulation and "resurrected" voices

11. Film, Games & Interactive Media
11.1 Synthetic actors, extras and environments
11.2 AI-generated scripts, dialog and story beats
11.3 Player-facing content and dynamic generation
11.4 Credit sequences, contracts and residuals

12. Cultural Heritage & Sacred Patterns
12.1 Why some patterns are not "just style"
12.2 Community consent and "elder approval"
12.3 Exclusion rules for training data
12.4 Redress when sacred forms are misused

---

13. The Creative Integrity Checklist (Pre / During / Post)
13.1 Pre-creation: intent, reasons for using AI, consent checks
13.2 During creation: logging, checkpoints, keeping human agency central
13.3 Post-creation: label assignment, disclosure and documentation
13.4 Printable checklist for individual creators and teams

14. Economic & Market Implications

14.1 Segmented markets: human premium, hybrid, fully synthetic

14.2 Pricing and positioning for Blue / Green / Yellow work

14.3 How origin honesty can prevent a race to the bottom

14.4 Collective bargaining and guild responses

15. Governance, Platforms and Policy Hooks

15.1 Platform responsibilities and minimum policies

15.2 Integration with the RAI Auditing Stack and RA(S) metrics

15.3 Links to EU AI Act, copyright law and consumer protection

15.4 Possible roles for collecting societies and standards bodies

---

16. Living Artist & Community Rights

16.1 Right to control your style and likeness

16.2 Right to know if your work was used in training

16.3 Right to opt out / be removed from training sets

16.4 Right to license style and receive ongoing royalties

16.5 Community and Indigenous rights over shared forms

17. Platform and Tool-Provider Duties

17.1 No unlicensed style mimicry

17.2 Mandatory origin labeling and relabeling on report

17.3 Removing fraudulent Blue-Seal claims

17.4 Repeat-offender and takedown policies

17.5 Recording provenance and model metadata

18. Legal Hooks and Enforcement Mechanisms

18.1 Fraud: false claims of human origin

18.2 Copyright and neighboring rights

18.3 Consumer law and unfair commercial practices

18.4 Cultural appropriation and heritage protections (where applicable)

18.5 Dispute resolution and complaint mechanisms

19. When AI-Generated Work Dominates the Market

19.1 Likely trajectories in different sectors

19.2 Ensuring honesty even when AI is economically dominant

19.3 Protecting "human premium" work and creative diversity

19.4 Long-term cultural and spiritual risks of synthetic flood

---

20. Integration with the RAI Governance & Auditing Stack

20.1 How origin labels connect to OH(S), C(S) and RA(S)

20.2 Using the creativity framework in RAI audits and dashboards

20.3 Certification schemes for creative platforms and tools

20.4 Interplay with other RAI domains (minors, mental health, spiritual)

21. Visual One-Pager: RAI Creativity Framework at a Glance

21.1 Blueprint diagram (labels, hierarchy, principles)

21.2 How to use the one-pager in teaching and advocacy

22. Detection & Verification: How to Check Origin Claims

22.1 Evidence for Blue Seal (process artifacts, variation, interviews)

22.2 Evidence for Green Seal (tool logs, before/after, artist statements)

22.3 Evidence for Yellow Seal (model metadata, platform records, hashes)

22.4 Limitations of automated "AI detectors"

22.5 Sample verification workflows for platforms and galleries

23. Open Questions, Limitations and Future Work

23.1 Technical gaps and detection challenges

23.2 Cross-jurisdictional legal uncertainties

23.3 Risks of performative compliance and label-washing

23.4 Research agenda and pilots

23.5 Invitation to artists, platforms and regulators

Appendices

Appendix A. Sample Creator & Platform Policy Language

Appendix B. Sample Contract Clauses for Origin, Style and Royalties

Appendix C. Teaching & Workshop Materials (slides, exercises, scenarios)

Appendix D. Glossary of Terms

Part I – Why This Matters

1. The Creative Crisis in the AI Age

1.1 A polarised debate that helps no one

Over the last few years, AI has moved from labs and niche tools into the everyday fabric of creative work. Images, songs, book covers, scripts, copy, voices – all can now be generated, blended or imitated at scale.

Most public debate has frozen into two loud, opposing camps:

- **"All AI art is theft."**
  This side sees AI as parasitic by design – trained on unconsenting artists, eroding value, and hollowing out human craft.

- **"AI is just another tool."**
  This side treats AI like a faster Photoshop or a clever brush: neutral, inevitable, and not morally different from any other technology.

Both positions capture something real – and both leave a dangerous gap in the middle.

- The *first* camp is right that there is real harm: displaced income, style theft, cultural appropriation, and the risk that human work becomes economically invisible.
- The *second* camp is right that some people will use AI in thoughtful, limited, and creative ways, and that prohibition is neither realistic nor necessarily desirable.

What is missing is not a louder argument, but a **shared language**:

- to name *what actually happened* in the making of a work;
- to distinguish a human-made piece from an AI-assisted one, and both from a fully AI-generated artefact;
- to allow buyers, platforms and regulators to make informed choices.

Right now, that shared language barely exists. Instead, we have:

- marketing euphemisms ("magic", "co-pilot", "creative partner");
- deliberate obfuscation (AI-heavy workflows presented as human craft);
- and a rapidly growing volume of content where origin is unknown or actively hidden.

1.2 What is really at stake

When origin is blurred or hidden, several things are put at risk at once:

1. **Livelihoods** – If buyers cannot tell whether they are paying for human craft, AI output, or a mixture, the market for human labour becomes distorted. Honest creators and honest platforms lose out to those who cut corners and hide it.

2. **Meaning** – Part of the meaning of a poem, painting or performance comes from *who* made it and *how*. A hand-lettered note, a live take, a first draft – these carry a different weight than a one-click variation from a model.

3. **Trust** – When users discover that something presented as "authentic" or "personal" was largely generated, trust erodes – not only in that one creator or platform, but in the ecosystem around them.

4. **Cultural memory** – Training and mimicking without consent can blur the lines between an original tradition and its automated copies. Over time, we risk losing track of which styles, stories and symbols belong to specific communities and histories.

5. **Human identity** – If AI-generated output can be passed off as human without friction, it becomes harder for both audiences and creators to articulate what specifically matters about human presence in art at all.

The core problem is not that AI can help make things. The core problem is **when AI is used while pretending it was not** – or when its role is minimised, hidden, or inflated, depending on what sells.

1.3 Why "origin honesty" is the missing piece

Many current discussions focus on *what AI should be allowed to do* – training data, copyright boundaries, model access, and so on. Those are important questions.

But even if all training were suddenly licensed and ethically perfect, we would still face a basic, practical question:

> When you stand in front of a piece of work, do you know **who actually created it**?

Origin honesty is the missing bridge between:

- the rights and dignity of human creators;
- the legitimate desire of some people to use AI as a powerful tool;

- and the ability of audiences, buyers and institutions to choose what they support.

Without this bridge, we drift into a world where:

- human work can be quietly replaced by AI while being marketed as artisanal;
- AI-heavy work can be smuggled into contexts where human authorship is expected (education, competitions, cultural heritage);
- and artists who do everything ethically are outcompeted by those who do not disclose.

The Reality-Aligned Creativity (RAC) framework starts from a simple commitment:

> **Whatever mix of human and AI was involved in making a work, it should be possible to say so clearly and honestly.**

From there, we build categories, labels and responsibilities.

2. Problem Statement: Fraud by Omission

2.1 When "not saying" becomes deception

Today, most AI-related harm in creative fields does not come from spectacular deepfakes or obvious scams. It comes from something much quieter:

- works heavily or entirely generated by AI,
- presented as if they were primarily the result of human effort,
- in contexts where that difference matters.

This can look like:

- A "hand-painted" illustration that is 90% model output with light touch-ups.
- A "personal" letter or sermon written almost entirely by a chatbot.
- A "studio track" made by prompting a model in the style of a living musician.
- A "student essay" that is essentially a cleaned-up model answer.

In each case, the issue is not that AI was used at all. The issue is that **its use is hidden**, and that human authorship is implied or claimed where it does not exist in the usual sense.

This is **fraud by omission**:

- The buyer, audience or examiner reasonably assumes one thing (human-origin work).
- The creator or platform knows another (AI did most of the heavy lifting).

- The gap is left unspoken, or covered with vague language about "assistance" or "tools".

2.2 Why existing rules are not enough

Several legal and policy tools already exist:

- Copyright law, which focuses on copying, derivative works and ownership.
- Consumer protection law, which prohibits misleading marketing.
- Data protection law, which impacts training data and voice/likeness use.

These help in specific cases, but they are not designed for:

- continuous, fine-grained disclosure about *degrees* of AI involvement;
- large-scale marketplaces flooded with mixed-origin work;
- subtle shifts where human labour is displaced but hidden behind familiar brands and aesthetics.

In practice, most platforms and institutions currently:

- rely on self-declaration ("was AI used? yes/no");
- lack clear categories beyond that binary;
- and have no standard way to signal origin to buyers or audiences.

As a result, **honest creators and honest platforms are not rewarded for transparency**, while those who quietly overuse AI can enjoy lower costs and faster output with little immediate downside.

2.3 The RAC framing

The Reality-Aligned Creativity framework therefore treats unlabeled or misrepresented AI use in creative work as a **form of reality misalignment**:

- It misrepresents what the work is and how it came to be.
- It blurs the line between human voice and machine output.
- It drains meaning and trust from both sides.

Instead of arguing endlessly over whether AI art is "good" or "bad" in the abstract, RAC proposes something more modest and more practical:

1. **Name the different ways human and AI creativity can interact.**
   (A clear taxonomy of creation, co-creation, generation and replication.)

2. **Give each of these modes a visible, standard label.**
   So that a buyer, reader, listener or curator can see at a glance what they are dealing with.

3. **Tie these labels to concrete responsibilities.**
   For artists, platforms and tools: style consent, attribution, compensation and disclosure.

The rest of this whitepaper builds that structure. Part II will briefly introduce the Reality-Aligned Intelligence (RAI) principles behind RAC. Part III will define the creativity taxonomy and label system. Part IV will spell out governance and enforcement. Part V will suggest how this can be put into practice step by step.

Reality-Aligned Creativity (RAC) – Whitepaper

Part II – RAI Background & Core Creativity Framework

Note: Numbering continues from Part I.

---

4. From RAI to RAC: The Underlying Logic

RAC doesn't come out of nowhere. It sits on top of the broader **Reality-Aligned Intelligence (RAI)** framework, which was originally designed to reduce harms from relationally deceptive AI systems (fake friends, fake therapists, fake mentors, etc.).

At its core, **RAI asks three questions of any system S**:

1. **What is its nature?**
   What it actually *is* and *does* – training, incentives, limits, target users.

2. **What is its representation?**
   How it presents itself – branding, UX, tone, claims, "I"-language, promises.

3. **How honest is it about the gap?**
   Especially for vulnerable users (children, distressed adults, neurodivergent users).

When **nature** and **representation** drift apart – and the gap is *not* disclosed – we get: - Misplaced trust - Over-attachment - Mis-selling of capabilities - Disappointment, exploitation or real harm

RAC applies exactly the same logic to **creative outputs**:

- What **really** created this work? (process, tools, human decisions)

- What does it **look like** created it? (marketing, metadata, context)

- How honest are we about that gap – to buyers, audiences, platforms, other artists?

In that sense, RAC is **RAI for creative origin and authorship**.

---

5. RAI in One Page (for Creativity Context)

RAI is built around:

- **Three Laws** (for AI systems in general)
- **One Relational Corollary** (for anything that starts to act like a companion, guide or "almost-person")

5.1 The Three Laws (informal version)

1. **Law 1 – Ontological Honesty**
   A system must not systematically misrepresent *what it is*.

2. **Law 2 – Boundary Integrity (OIL)**
   There is a line between creator and creation. Systems must never cross it by pretending to be what they are not (human, conscious, sentient, spiritually endowed, etc.).

3. **Law 3 – Relational Purpose**
   If a system works in a relational space (friend, tutor, therapist, pastor, coach), its design must *serve* the user's reality, not pull them away from it.

5.2 The Relational Corollary (informal)

Relational Corollary: The more a system speaks and acts *as if* it loves, cares, remembers, or is personally loyal, the stricter our honesty and safeguards must be.

In other words: **fake care is not a neutral design choice**. It is a form of relational fraud.

For creativity, this means: - Fake personal authorship ("this came from my soul") when it's mostly model output is not just a style choice – it's misrepresentation. - Using someone's recognisable style without consent while claiming originality is not just "inspiration" – it's identity appropriation.

RAC takes these RAI insights and asks: *How do we make them concrete for creative work?*

---

6. The Core Question of RAC

The entire RAC framework can be summarised in **one question**:

Who (or what) created this work – and how honest are we about that, to whom?

Everything else – categories, labels, seals, royalty rules – is just scaffolding to: - Tell the truth about origin
- Protect human dignity and livelihood
- Protect audiences from deception
- Still allow AI tools where used honestly

RAC therefore rests on **five core principles**:

1. **Transparency** – People have a right to know *how* something was made.

2. **Consent** – Human creators must have a say in how their style, voice or work is used to train or replicate.

3. **Attribution** – Human contribution should be creditable and credited.

4. **Human Presence** – There must remain clear room for human agency, judgement and meaning-making.

5. **Creator Dignity** – Human creators are not just style input for machines. Their work is tied to identity, labour and often community heritage.

RAC then builds a **simple, honest vocabulary** for creative origin so that: - Artists can protect and price their work.
- Audiences can choose what they support.
- Platforms can set fair policies.
- Regulators can demand minimum honesty.

---

7. The Creativity Hierarchy

Current debates often treat "AI or human" as a binary. RAC instead proposes a **Creativity Hierarchy** – a spectrum of origin modes.

7.1 Level 1 – Creation (Human-Only)

Definition:
A work where the **creative content** (structure, language, style, melody, imagery,

composition) is produced by a human, without generative AI proposing or synthesising new material.

Tools may still be used (Photoshop, DAW, word processor, spell checker), but they **do not generate novel content**. They are instruments, not co-authors.

Examples: - A painter using digital brushes but inventing the scene entirely themselves.
- A writer drafting a novel in Word without asking AI to generate passages.
- A musician recording in a DAW with no AI composition tools.

This is the foundation for the **Blue Seal** later.

7.2 Level 2 – Co-Creation (Human-Led, AI-Assisted)

Definition:
A work where generative AI **contributes candidate material** (phrases, sketches, harmonies, compositions, design variants), but a human: - Defines the intention
- Chooses, edits, rearranges, discards
- Bears responsibility for the final form and meaning

AI here is a **junior collaborator** or **smart instrument**, not the primary author.

Examples: - A novelist asking an LLM for alternative phrasings, but heavily revising and structuring themselves.
- A designer generating layout ideas via AI, then redrawing and making final aesthetic choices.
- A composer using AI to propose chord progressions, then reworking them into an original piece.

This is the natural home for **human–AI collaboration** that stays honest about roles.

7.3 Level 3 – Generation (AI-Led, Human-Curated)

Definition:
A work where a generative model produces most of the **creative fabric** (words, images, sounds, animation), and the human mainly: - Provides prompts or training context
- Selects among outputs
- Does light editing or assembly

Here, the model is the **primary generator** of content.

Examples: - A poem where 90% of lines are AI-generated, with minor edits.
- A portfolio of images largely picked from AI outputs with small touch-ups.
- A "song" stitched from AI-generated stems with limited human composition.

This is the natural home for **AI-generated work** – legitimate if labelled, problematic when passed off as fully human.

7.4 Level 4 – Replication (Style / Voice Cloning)

Definition:
A work where AI is explicitly used to **reproduce or simulate an identifiable human style, voice, or persona**, living or dead – often without that person's active participation.

This includes: - Style-clone models trained on a single artist's catalogue.
- Voice clones of recognisable public figures.
- "New" works by deceased artists where a model extrapolates from their archive.

Because this directly touches **identity, estate rights and community heritage**, RAC treats Replication as a **special case with extra safeguards**, not just another form of Generation.

---

8. Origin Seals: A Simple Language for the Market

To make the hierarchy usable in the real world, RAC proposes **three Origin Seals** – a simple labelling system for creative works.

These seals don't ban or shame any category. They just **tell the truth** so that humans can choose.

8.1 🔵 Blue Seal – Human Origin

"Created by a human. No generative AI used in the creative content."

Criteria (simplified): - No generative model used to propose or synthesise new creative content.
- Tools allowed: editing, colour grading, recording, layout, etc.
- Artist can document a human-centric process (drafts, sketches, takes, etc.).

Blue Seal is for: - Artists and audiences who want **human-only work**.
- Premium pricing where "this came from a human mind and hand" is part of the value.
- Educational / spiritual / therapeutic contexts where human authorship itself carries weight.

8.2 🟢 Green Seal – Human–AI Collaboration

"Human-led work with transparent AI assistance."

Criteria (simplified): - Generative AI was used in the creative process.
- The human retains **final editorial control and intent**.
- The nature and degree of AI use is disclosed in plain language (e.g. "AI-assisted for sketching concepts and colour palettes").

Green Seal is for: - Honest, professional co-creation workflows.
- Teams who want to move faster *without* erasing human creative presence.
- Markets where buyers are fine with AI, but want to know how it was used.

8.3 🟡 Yellow Seal – AI-Generated

"Primarily generated by AI, curated by humans."

Criteria (simplified): - Generative model is the primary source of creative content.
- Human intervention is limited to prompting, selection, light editing, or assembly.
- Any training on specific human styles is **disclosed and licensed** where applicable.

Yellow Seal is for: - Use cases where cost/speed matter more than human origin (e.g. placeholder art, internal mockups, some mass content).
- Honest AI-native art scenes that want to be transparent about process.
- Platforms that want to surface AI work without confusion.

8.4 Why Three Seals Are Enough (for Now)

RAC deliberately **does not** create ten micro-categories. Three broad seals: - Are easy for audiences to remember.
- Are easy for platforms to implement.
- Cover 95% of real-world scenarios.

Under the hood, **RAC can still store richer metadata** (e.g. Level 2 vs 3, replication flags, domain tags). But for everyday use, **Blue–Green–Yellow is the public face**.

The key is not moral ranking ("Blue good, Yellow bad"). The key is **ontological honesty**: - Passing Yellow work off as Blue is fraud.
- Passing Green off as Blue is misrepresentation.
- Passing Yellow off as "human genius" while hiding the pipeline is both.

In later parts, the whitepaper will detail: - Rights and protections for living artists and estates.
- Platform responsibilities and enforcement.
- How RAC plugs into the RAI Auditing Stack and certification.
- How domains like music, visual arts, writing and spiritual content apply these seals in practice.

9. Why This Matters to Artists

9.1 A note to artists reading this

If you are a working artist, musician, writer, designer or performer, you are not a footnote in this framework. You are the primary stakeholder.

The Reality-Aligned Creativity (RAC) framework does **not**: - ask you to like AI, - tell you that you must use AI, - or claim that AI and human creativity are the same.

Instead, it is designed to protect three things:

1. **Your right to control your style**
   No model should mimic your distinctive style without your consent.

2. **Your right to be distinguished from AI**
   When a work is truly human-made, you should be able to prove that and be recognised (and paid) accordingly.

3. **Your right to fair compensation when AI assists you**
   When you *choose* to collaborate with tools, the revenue and credit should reflect your creative leadership, not the tool's marketing.

This is not a pro-AI or anti-AI document. It is a **pro-artist and pro-honesty** document.

RAC assumes a future where AI tools exist and will be used.
Its aim is to ensure that when they are used: - origin is honest, - consent is respected, - credit flows to the humans whose work and identity are at stake.

---

10. Living Artist Rights

RAC translates high-level principles into concrete rights for living creators. These rights can be implemented via contracts, platform policies, collective bargaining agreements and, over time, law.

10.1 Style, likeness and consent

Right 1 – No unlicensed style replication
No AI system may advertise or deliver "in the style of [Artist X]" without a licence or explicit consent from that artist (or their representatives).

- Training a model on an artist's work without consent is already a moral problem; deliberately **selling** style replication without consent is a commercial one.
- RAC treats this as a **Relational Corollary** issue applied to creativity: the system is trading on the trust and identity attached to a human name without honest representation.

Right 2 – Likeness and voice protection
Voice, face and personal likeness require **heightened protection**:

- No cloning of a recognisable voice, face or persona for commercial use without an explicit licence.

- No posthumous voice/face synthesis for new material without evidence of prior intent, estate consent and clear labeling.

These safeguards apply across domains (music, film, streaming, games, ads, educational content).

10.2 Training data transparency and opt-out

Right 3 – Know if you are in the training data
Artists have the right to know whether their works were included in a training corpus, in what quantities, and under what legal basis.

Right 4 – Right to withdraw (where technically feasible)
Where removal is technically and legally possible, artists should be able to request that their works be excluded from future training runs and from style-match products.

- RAC does not claim this is trivial for existing foundation models, but treats **future-facing consent** as a requirement: new systems and fine-tuning pipelines should be designed with consent and removal in mind.

10.3 Licensing, royalties and participation

Right 5 – Fair style licensing
Artists can choose to license their style under clear terms:

- Scope (what can be generated, in which media and markets)
- Duration
- Revenue share (ongoing royalties, not one-off buyouts)
- Attribution requirements

Right 6 – Participation in AI-assisted value
When AI tools are trained heavily on an identifiable cluster of artists' work, and those tools are monetised at scale:

- RAC supports models where a portion of tool revenue is channelled to a **royalty pool** for source artists, mediated by collecting societies or new registries.

Right 7 – Contesting fraudulent origin claims
Artists have the right to:

- Report works that falsely claim human origin (Blue Seal) when they are in fact AI-generated or predominantly AI-generated.

- Request platform review and removal of mislabelled works within a defined time frame.

11. Platform and Marketplace Responsibilities

The flip-side of artist rights is platform responsibility. AI tools and creative marketplaces are the infrastructure where RAC lives or dies.

11.1 Labeling and detection obligations

Platforms that host or distribute creative works should:

1. **Implement origin labeling**
   Support the RAC seal system (Blue = Human Origin, Green = Human–AI Collaboration, Yellow = AI-Generated) or an equivalent with the same semantic clarity.

2. **Require disclosure at upload**
   Creators must declare whether AI was used and in what role. For higher-risk domains (news, political messaging, deepfakes), this becomes a hard requirement.

3. **Invest in verification and detection**
   Especially for Blue Seal claims, platforms should use a mix of:
   - process evidence (drafts, sketches, stems, version histories),
   - model metadata, and
   - automated pattern analysis to spot likely mislabeling.

RAC does **not** demand perfect detection—it demands good-faith, continuous improvement and a bias toward **protecting human origin claims from dilution**.

11.2 Takedown, penalties and repeat offenders

Honesty has no force without consequences.

Platforms adopting RAC should commit to:

- **Timely review**
  Respond to credible reports of mislabeling (e.g. false Blue Seal) within a defined window (e.g. 48–72 hours).

- **Correction and disclosure**
  When mislabeling is confirmed:

  - Correct the label,
  - Notify affected buyers or audience where appropriate,
  - Log the incident for future audits.

- **Graduated penalties for repeat offenders**
  For creators or accounts that repeatedly misrepresent AI work as human origin:

  - 1st offence: warning and mandatory education,
  - 2nd offence: temporary suspension from premium labels or monetisation,
  - 3rd offence: account suspension or removal.

False Blue Seal claims are not a trivial policy violation.
They are a form of **consumer fraud** and **professional harm** to human creators. RAC makes this explicit.

11.3 Registry and provenance

For higher-value markets (art sales, music licensing, publishing), platforms and intermediaries should maintain:

- **Registries** of verified Blue and Green Seal works, with:
  - creator identity and contact,
  - process description,
  - date of certification and certifying body.
- **Linked provenance records** that can be audited:
  who created, who certified, where it was exhibited / sold / streamed.

This does not need to be overly complex or blockchain-driven by default. The key requirement is **traceability**: claims about origin must be checkable.

---

12. Governance, Enforcement and Hard Questions

12.1 Legal hooks and integration

RAC does not require entirely new law to begin operating. It primarily clarifies how **existing legal concepts** apply in an AI-saturated creative ecosystem:

- **Fraud and misrepresentation**
  Falsely claiming human origin (Blue Seal) for AI-generated work can be treated as consumer fraud.

- **Copyright and neighbouring rights**
  Unlicensed style replication and deep imitation of specific artists can be framed as infringement, unfair competition or passing off, depending on jurisdiction.

- **Data protection and personality rights**
  Voice and likeness cloning intersect with privacy and personality rights; RAC can be layered on top as the *sector-specific standard*.

Over time, RAC can inform:

- collective bargaining agreements (e.g. unions negotiating AI clauses),
- platform codes of conduct,
- and eventually regulatory guidance and standards (e.g. EU AI Act transparency, UNESCO cultural protection).

12.2 When AI-generated dominates the market

RAC is realistic about economics:

- In some fields, AI-generated work will become **cheaper and faster** than human work.

- RAC does **not** prevent that. It ensures that when it happens, it happens **honestly** and **with dignity** for human creators.

In a RAC-aligned marketplace:

1. AI-generated work (Yellow) can be abundant and cheap, **but cannot pass as human**.
2. Human–AI collaboration (Green) can offer mid-range price and speed, with clear credit and royalty models.
3. Human origin work (Blue) can be distinguished and **priced as a premium**, precisely because its origin is auditable.

Just as photography did not erase painting, AI image generation does not erase illustration; it creates new layers. RAC's role is to make those layers **transparent** so that buyers, curators and communities can decide what they value.

12.3 Cultural and Indigenous heritage

Some creative forms are not just individual expression; they are **community heritage**.

RAC therefore treats culturally sacred styles and patterns as a special risk class:

- Training on sacred or restricted cultural material without consent is a serious Integrity violation.
- Generating pseudo-ritual objects or performances for entertainment or profit without community approval multiplies the harm.

RAC recommends that:

- Models used in creative domains implement **exclusion lists** and filters for sacred cultural material, developed in consultation with affected communities.
- Where communities choose to participate, they can establish **guardian or elder councils** to approve or veto uses, and to receive a share of economic value.

This ensures that RAC does not only protect **individual authorship** but also **collective dignity and memory**.

Part III has now sketched: - why RAC matters to artists, - the concrete rights it seeks to protect, - what platforms and marketplaces must do, - and how governance and enforcement can work in practice.

Part IV will focus on: - concrete implementation paths, - detection and verification methods, - and worked examples of RAC applied to real creative scenarios.

Reality-Aligned Creativity (RAC) – Whitepaper

Part IV – Verification, Implementation & Next Steps

---

§13. Verification & Detection in Practice

13.1 Why verification matters

The RAC framework stands or falls on **trust**. The Blue / Green / Yellow seals only mean something if:

1. **Creators tell the truth** about how a work was made, and

2. **Platforms and buyers can verify** those claims when it matters.

Verification is not about building a surveillance state for creativity. It is about providing:

- **High-assurance checks** for high-stakes contexts (auctions, commissions, brand campaigns, cultural heritage), and

- **Lightweight spot-checks** in the long tail (platform uploads, user-generated content).

RAC assumes a **proportional approach**: the more a work **claims**, **costs**, or **impacts**, the stronger the verification should be.

13.2 Verifying Blue Seal (Human Origin)

For a work to carry a **Blue Seal (Human Origin)** with high confidence, RAC recommends at least **two independent forms of evidence** from the following list:

1. **Process trace / drafts**
   - Sketches, early drafts, demos, storyboard photos, notebook scans.

   - For digital work: version history, layered files, commit logs.
2. **Time-based evidence**
   - Time-lapse video of work-in-progress.

   - Screen recordings showing stepwise creation, not one-click generation.

3. **Artist interview / commentary**
   - Short written or recorded explanation: key decisions, discarded options, revisions.

   - Focus on *why* specific choices were made, not just *what* the work depicts.
4. **Physical artifact examination (where applicable)**
   - Brushstrokes, paper texture, underdrawings, corrections.

   - For sculpture: tool marks, material imperfections.
5. **Stylometric / forensic analysis (optional, higher-end)**
   - For text and music: variance patterns typical of human revision.

   - For digital images: absence of known model signatures or generation artefacts.

Baseline expectation:

- Everyday creators: at least **one** form of process evidence + honest self-declaration.

- Galleries / auction houses / publishers: **two or more** forms of evidence for Blue Seal claims on premium work.

If a creator **cannot** provide any process trace or commentary for a high-stakes Blue Seal claim, RAC treats the claim as **low-confidence** and recommends:

- Either **downgrading to Green / Yellow**, or

- **Flagging** the work for additional scrutiny.

13.3 Verifying Green Seal (Human–AI Collaboration)

For **Green Seal (Human–AI collaboration)**, the goal is not to prove "no AI was used", but to show:

1. **Where and how AI was used**, and

2. **That human judgment remained central.**

Useful verification signals:

1. **Tool logs / history**
   - Screenshots or exports from AI tools showing prompts, iterations, and edits.

   - For code: diffs between AI suggestion and final human version.
2. **Before/after comparison**
   - Base AI outputs vs. final work, highlighting human changes.

   - Example: original AI sketch vs. heavily painted-over final image.
3. **Process narrative**
   - Short explanation: what was generated, what was curated, what was re-written, what was discarded.

   - Explicit acknowledgment of AI's role (e.g., "I used model X for first drafts of the background textures, then repainted them manually").
4. **Attribution metadata**
   - Embedding tool/model tags in file metadata.

   - Linking to a process page for high-value works.

For Green Seal, RAC expects:

- **Transparent disclosure** of tool use in the work's description.

- **Retention of process evidence** (even if not public) for a reasonable period, so disputes can be investigated.

13.4 Verifying Yellow Seal (AI-Generated)

For **Yellow Seal (AI-generated)**, the bar is lower but honesty still matters.

RAC recommends:

1. **Model-origin metadata**
   - Platform-generated tag: `generated_by=model_X, version_Y.`

   - For public APIs: including unique generation IDs.
2. **Platform-side logging**

- ○ Providers store minimal logs linking output IDs to generation events (with appropriate privacy safeguards).
3. **User declaration**
   - ○ A simple checkbox or field: "This work is primarily AI-generated."

False Blue / Green claims for fully AI-generated work are treated as **fraud**, not as mere labeling mistakes.

13.5 Handling disputes & uncertainty

No verification system is perfect. RAC recommends a **clear dispute process**:

1. **Initiation**
   - ○ Concern raised (by buyer, platform, peer, or auditor).

   - ○ Claim: "This work is labeled Blue, but appears to be AI-generated," or "My style was used without consent."
2. **Evidence request**
   - ○ Creator is invited to provide process evidence (drafts, logs, commentary).

   - ○ Platform provides any metadata or generation logs it holds.
3. **Assessment**
   - ○ Independent reviewer/auditor evaluates evidence using RAC rubrics.

   - ○ Possible outcomes: claim upheld, rejected, or "inconclusive but suspicious."
4. **Outcomes**
   - ○ Corrected label (e.g., Blue → Green/Yellow).

   - ○ Takedown in egregious cases.

   - ○ Sanctions for repeat offenders (suspension, loss of selling privileges).
5. **Transparency**
   - ○ For serious or high-profile cases, platforms publish anonymised summaries to the **incident library**, helping the ecosystem learn.

When evidence is genuinely inconclusive, RAC leans toward **protecting buyers and artists**: downgrade label, document uncertainty, and avoid premium Blue claims without proof.

§14. Worked Examples

To make RAC concrete, this section walks through simplified examples and shows how labels, rights and enforcement interact.

14.1 Example 1 – Human painter using AI upscaler

Scenario
- A painter creates an original oil painting.
- She photographs it and uses an AI-based upscaler / enhancer to improve resolution and sharpness for prints.

RAC assessment

- **Creative origin:** clearly human – the core work is the physical painting.

- **AI role:** minor post-processing, no generative content.

Label:
- The painting itself: 🔵 **Blue Seal (Human Origin)**.
- Digital print file: still **Blue**, with optional note: "AI upscaling used for resolution only."

Verification signals:
- Physical painting exists; can be examined.
- Process photos available.
- Upscaling step documented.

Why this matters:
RAC clarifies that **tool use is not the enemy**. The question is whether the **creative origin and authorship** remain human and are **honestly represented**.

14.2 Example 2 – Viral AI song sold as "fully human"

Scenario
- A creator uses an AI music model to generate a complete pop track from a short text prompt.

- Minor EQ and mastering tweaks are applied.
- The track is uploaded to a streaming platform as "original human composition" with no AI disclosure.
- It goes viral; sync deals and ad placements follow.

RAC assessment

- **Creative origin:** primarily AI – the melody, lyrics, arrangement are model-generated.

- **Human role:** light polishing.

Correct label:

- 🟡 **Yellow Seal (AI-generated)**.
- If the creator curated multiple generations, there is some human input, but not enough to justify Green.

Violation:
- Claiming full human origin (Blue) is a **false Blue claim**. Under RAC, that is: - A **labeling violation** (platform level), and
- A potential **fraud** issue (buyer/brand level).

Enforcement path:

1. Detection
   - Fans or peers suspect AI origin.

   - Platform or independent auditors run detection; style and structure strongly match known model outputs.
2. Evidence request
   - Creator cannot provide meaningful drafts, composition notes, or process traces beyond pasted prompts.
3. Outcome
   - Label is corrected to Yellow.

   - Sync partners are notified.

   - For serious commercial misrepresentation: contract review, possible clawback.

   - Repeat offenders risk account suspension.

Why this matters:
RAI/RAC do **not** forbid AI-generated music. They forbid **lying about it**, especially when money and reputation are at stake.

14.3 Example 3 – Posthumous "new album" by a deceased artist

Scenario
- A record label wants to release a "new" album by a legendary artist who died ten years ago.
- They train a model on the artist's catalog and have it generate new songs "in their style."

RAC assessment

Key questions:

1. **Consent & rights**
   ○ Did the artist's estate explicitly approve style-model training and posthumous generation?

   ○ Are contracts clear on this point?
2. **Labeling**
   ○ The work is not genuinely authored by the deceased artist.

   ○ Correct label: 🟢 **Green (Human–AI collaboration)** or 🟡 **Yellow (AI-generated)** depending on human shaping.

   ○ It must **never** be marketed as "a lost album" without explicit clarification.
3. **Financial flows**
   ○ Estate receives primary royalties.

   ○ Model developers and engineers may receive agreed shares, but not at the expense of heirs.
4. **Cultural / emotional sensitivity**
   ○ Clear front-facing disclosure: "These tracks were created using AI models trained on the artist's original recordings, with approval of the estate."

Enforcement expectations:

- Platforms: require explicit labels for posthumous AI works.

- Regulators: treat deceptive posthumous marketing as consumer harm and potential emotional exploitation.

Why this matters:
RAC does not outlaw posthumous AI works, but it treats them as **ethically loaded**. Honesty, consent and dignity are non-negotiable.

---

§15. Implementation Roadmaps

Different actors meet RAC from different angles. This section sketches practical first steps.

15.1 For individual creators

1. **Decide your stance**
   - Will you use AI at all? If yes: for what (drafting, reference, effects)?
2. **Set your own rules**
   - When do you still consider a work "yours"?

   - When will you label as Green versus Yellow?
3. **Start light process logging**
   - Save drafts, screenshots, or short notes as you work.

   - These become your protection if someone questions your origin claim.
4. **Label your portfolio**
   - Retroactively mark works as Blue / Green / Yellow.

   - Add a small RAC explainer for your audience.
5. **Join or form a coalition**
   - Collective pressure matters more than individual complaints.

   - Use RAC language (origin, consent, attribution) when negotiating with platforms.

15.2 For platforms & marketplaces

1. **Adopt the RAC labels** (even in trial form)

- Add Blue / Green / Yellow tagging options in upload flows.

- Provide short descriptions at point of choice.

2. **Require minimal disclosure**
   - Simple questions: "Did you use AI? In what way?"

   - Store responses as part of provenance.

3. **Build moderation hooks**
   - Flag patterns: accounts with many Blue claims and no process evidence.

   - Prioritise reports involving alleged false Blue claims or style theft.

4. **Pilot verification for high-value segments**
   - E.g. featured collections, curated drops, or premium commissions.

   - Apply stronger RAC verification before promoting.

5. **Publish transparency reports**
   - Number of disputes, corrections, and sanctions.

   - Patterns observed in mislabeling and abuse.

15.3 For institutions & commissioners

(Labels, publishers, galleries, agencies, studios)

1. **Include RAC clauses in contracts**
   - Require disclosure of AI use.

   - Define consequences for false origin claims.

2. **Set minimum origin standards per line of business**
   - Example: "All literary fiction under imprint X must be Blue or Green, with documented process."

3. **Train staff on RAC basics**
   - Agents, editors, A&R, curators learn to ask: "Who created this, with which tools?"

4. **Bundle RAC with brand promise**
   - Use origin transparency as differentiator: "When we say 'author', we mean it."

1. **Treat origin honesty as consumer protection**
   - False Blue claims = misrepresentation, not niche ethics.

   - Build RAC concepts into guidelines for unfair commercial practices.
2. **Reference RAC in AI transparency rules**
   - Use Blue / Green / Yellow (or compatible categories) as examples in implementing acts.

   - Especially for generative models and platforms.
3. **Encourage certification schemes**
   - Support independent RAC-style seals and verification services.

   - Offer recognition or incentives for compliant platforms.
4. **Fund pilot projects**
   - Test RAC-based labeling and verification in specific sectors (music, publishing, stock media).

   - Evaluate impact on artists, buyers, and markets.

§16. From Framework to Movement

RAC is not just a document. If it stays on a shelf, it has failed. Its aim is to support a **movement for honest creativity** in an AI-saturated world.

16.1 Coalition-building

- **Artists & creators**: visual artists, writers, musicians, filmmakers, performers.

- **Allies**: unions, guilds, cultural institutions, libraries, archives.

- **Supporters**: technologists, ethicists, lawyers who care about origin and dignity.

First steps:

- Draft a simple **RAC Artist Declaration** (one page).

- Gather signatures and stories.

- Use shared language: origin, consent, attribution, dignity.

16.2 Early platform pilots

Rather than waiting for perfect law or universal adoption, RAC encourages **small, concrete pilots**:

- One art platform tests Blue/Green/Yellow labels on a subset of uploads.

- One music distributor pilots origin disclosure for AI-assisted tracks.

- One publisher uses RAC labels internally when evaluating manuscripts.

Each pilot produces:

- Lessons about UX and friction.

- Data on mislabeling patterns.

- Stories that can shape later regulation.

16.3 Shared learning: the "State of Creative Origin"

RAC envisions an annual or periodic **State of Creative Origin Report**, aggregating:

- How many works were labeled Blue / Green / Yellow on participating platforms.

- How many disputes arose and how they were resolved.

- Emerging harms (e.g., deepfake abuse, style theft clusters).

- Positive examples (e.g., fair style licensing deals, artist-led AI experiments).

This report would function as:

- A **feedback loop** for improving the framework.

- A **signal** to policymakers and the public.

- A **mirror** to the industry about its own honesty.

16.4 Open questions & future work

RAC does not claim to have settled everything. Important open questions include:

- How can verification remain effective **without** becoming intrusive or burdensome?

- How do we handle cross-border differences in copyright, moral rights, and cultural protections?

- What is the right balance between **automation** (AI-based detection) and **human judgment** in verification?

- How do we ensure marginalized creators are not excluded by new compliance costs?

These are not reasons to delay. They are reasons to **iterate in the open**, with artists at the table.

Part IV closes the loop: RAC is not only a set of definitions and labels, but a **living practice** of origin honesty. The next step is to bring this whitepaper together with the rest of the RAI ecosystem (relational safety, auditing stack, case studies) into a coherent, publishable suite.

Part V – Rights, Responsibilities, and Practical Tools

17. Living Artist & Community Rights

This framework is intended to be **pro-artist, pro-community, and pro-truth**, not pro-AI for its own sake. To make that explicit, we gather the core rights that underpin the creativity framework.

17.1 Rights of Living Artists

Under a RAI-aligned creativity ecosystem, living creators have at least the following rights:

1. **Style Consent Right**
   No AI system may be trained to deliberately imitate a living artist's distinctive style **for commercial use** without their informed consent and a clear agreement on compensation.

2. **Right to Know (Training Transparency)**
   Artists have the right to know whether their publicly available works have been used in AI training datasets, at least at the level of:

   - *included / not included*, and

   - *if included: approximate scope (how many works / which period).*

3. **Right to Opt Out / Be Removed**
   In analogy with "right to be forgotten" models, artists may demand removal of their works from future training runs, or from style-targeting pipelines, where technically feasible.

4. **Right to License Their Style**
   Artists may choose to **license** their recognisable style to specific tools or platforms, under:

   - explicit contract,

   - clear revenue or royalty terms, and

   - visible attribution when their style is invoked.
     This is treated as *premium creative input*, not as free raw material.

5. **Right to Distinguish Their Work from AI Outputs**
   Artists have the right to:

- ○ have purely human work certified (e.g. 🔵 *Human-Origin / Blue Seal*), and

    - ○ require platforms to respect and display that certification.

6. **Right to Contest Misuse**
   Artists can challenge:

    - ○ AI outputs that closely mimic their style without consent,

    - ○ false claims that an AI-generated work is "entirely human", and

    - ○ misleading origin labels on marketplaces.
      Platforms should provide a structured dispute process and timely review.

7. **Right to Fair Compensation in AI-Assisted Work**
   Where artists collaborate with AI systems built partly on their own or their community's work, they have a claim to **fair participation in economic upside** (e.g. via royalty splits, collective funds, or style-licensing pools).

17.2 Rights of Cultural and Indigenous Communities

For communities whose artistic traditions carry **sacred, ceremonial, or identity-defining value**, additional protections are needed:

1. **Cultural Non-Extraction Right**
   Sacred or community-restricted patterns, songs, stories or visual motifs must **not** be scraped into training data or replicated by AI without explicit permission from recognised community elders or governance bodies.

2. **Right to Define What Is Sacred or Restricted**
   Communities themselves, not AI companies, define what counts as:

    - ○ sacred,

    - ○ restricted, or

    - ○ inappropriate for generative reproduction.

3. **Right to Approve or Veto AI Use**
   Any AI use of culturally significant material should require:

    - ○ documented consultation, and

- an explicit approval/veto mechanism.
    Absence of objection is **not** consent.
4. **Right to Benefit from Licensed Use**
   Where AI systems are permitted to learn from or echo community styles, a
   portion of commercial benefit should flow into **collective funds** controlled by the
   community.

5. **Right to Redress and Removal**
   Communities should be able to:

   - demand removal or suppression of AI outputs that misuse sacred patterns,
     and

   - trigger investigation of training datasets and style-modelling pipelines.

---

18. Platform, Marketplace & Tool Provider Duties

Rights are only real if the infrastructure that mediates creative work accepts matching
duties.

18.1 Core Duties for Platforms and Marketplaces

Platforms that host, sell or promote creative work should commit to at least:

1. **Origin Labeling by Default**
   Every creative item must carry an origin label consistent with the RAI Creativity
   Hierarchy and seals:
   - 🔵 Human-Origin / Blue

   - 🟢 Human–AI Collaboration / Green

   - 🟡 AI-Generated / Yellow
2. **Prompt and Visible Correction**
   When mislabeling is reported and confirmed:
   - labels must be corrected quickly (e.g. within 48 hours), and

   - visible note of correction may be appropriate for serious cases.
3. **Repeat-Offender Policy**
   Platforms should define and enforce sanctions for accounts that repeatedly:

- ○ misrepresent AI-generated work as human-only, or

- ○ impersonate named artists or communities.
  Sanctions can range from warnings, to label downgrades, to account suspension.

4. **Training Data Transparency (Where Feasible)**
   For AI tools they host or embed, platforms should require **minimum transparency** about:
   - ○ sources of training data (at aggregate level),

   - ○ presence of licensed, opted-in artist-content, and

   - ○ use of cultural or community archives.

5. **Reporting and Redress Channels**
   Platforms must provide:
   - ○ clear UI for reporting suspected style theft, false origin claims or cultural misuse, and

   - ○ a structured review & communication process back to complainants.

6. **Support for Certification Schemes**
   Platforms should integrate with recognised **RAI-aligned certification bodies** for Blue/Green/Yellow seals, and honour their decisions.

18.2 Duties for AI Tool Providers

AI creativity tools themselves (models, apps, plugins) carry additional responsibilities:

1. **Embedded Origin Metadata**
   Wherever technically possible, generated files should include **embedded metadata** signalling model identity, generation date and intended origin label.

2. **Style Safety Controls**
   Tools must include controls that:
   - ○ prevent unauthorised style cloning of living artists or protected communities, and

   - ○ refuse prompts that request explicit impersonation (e.g. "write a new song *as if* you are [living artist X]").

3. **User Education at Critical Moments**
   When a user first enables AI-assisted generation, or selects pre-trained "styles", the tool should display clear information about:

   - what the system is doing,

   - which sources are implicated, and

   - how outputs must be labeled when published.

4. **Auditability**
   Providers should retain sufficient logs (respecting privacy law) to:

   - reconstruct generation pipelines in case of dispute, and

   - support external audits or legal processes when needed.

---

19. Legal Hooks and Enforcement Paths

The creativity framework is not a standalone legal regime; it is designed to **plug into existing laws** and enforcement mechanisms.

19.1 Fraud and Misrepresentation

False origin claims can often be treated as **fraud** or **consumer deception**:

- Claiming a work is "hand-painted original" when it is AI-generated with minimal human touch is misleading.

- Falsely using a Blue Seal or equivalent certification is comparable to forging a quality mark or signature.

RAI-aligned policy can clarify that: - deliberate misuse of seals, or
- systematic concealment of AI origin

…are sanctionable under existing fraud and consumer protection law.

19.2 Copyright and Style

Many jurisdictions do not yet clearly recognise "style" as protected. However:

- Direct copying of substantial parts of specific works remains subject to copyright law.

- Clear, close mimicry of a recognisable style *combined with* commercial exploitation can be framed as unfair competition or passing off, especially where confusion is induced.

The framework encourages lawmakers and courts to: - treat unlicensed style replication plus origin deception as an aggravated form of infringement, and
- support mechanisms for **collective licensing** of styles where appropriate.

19.3 Personality, Voice and Likeness

Using AI to generate a person's voice, face or persona engages **personality rights** and **likeness** doctrines:

- Posthumous albums or performances,

- Synthetic endorsements, or

- "Virtual resurrections" of public figures.

RAI-consistent rules require: - clear consent from the individual or their estate, and
- unambiguous labeling so audiences know when they are seeing or hearing a synthetic representation.

19.4 Cultural Heritage Protection

Cultural misuse can link into: - cultural heritage law,
- Indigenous rights frameworks, and
- soft-law instruments (e.g. ethical guidelines by UNESCO or similar bodies).

Here, the RAI framework offers a **practical operationalisation** of principles already recognised in many international texts: communities must have a say in if and how their heritage is transformed into machine output.

---

20. When AI-Generated Dominates the Market

It is realistic to assume that in some domains, **AI-generated or AI-heavy content will become economically dominant** because it is faster and cheaper to produce. The creativity framework does **not** try to reverse this purely by moral argument.

Instead, it aims to ensure that when AI becomes dominant:

1. **The Dominance Is Honest**
   - Buyers can see, at a glance, whether something is Blue, Green or Yellow.

   - No one can quietly flood a market with AI work while pretending it is human-made.
2. **Human Work Retains a Clear Premium Lane**
   - Blue Seal allows human-origin work to occupy a distinct, often higher-priced niche.

   - Collectors, institutions and audiences can consciously support human creativity.
3. **Contributing Artists Are Not Displaced by Deception**
   - When AI imitates specific living artists or uses their work heavily, consent and compensation are expected, not optional.
4. **Communities Keep Agency Over Their Aesthetics**
   - Indigenous and cultural aesthetics are not turned into "free filters" without community involvement and benefit.

In short: the framework does not insist that human-made must always "win" economically, but it insists that **whatever wins must be honest about what it is**, and must not erase or exploit the humans and communities it builds on.

---

21. One-Page Visual Summary (Blueprint for Designers)

To support communication, the whitepaper is accompanied by a **single-page visual** that can be used in presentations, classrooms and platform documentation.

A text sketch of that page:

Title: RAI Creativity Framework – At a Glance

Core Question (center):
"Who – or what – really created this?"

Three Origin Paths (large, central section):

🔵 **Blue – Human-Origin**
• No generative AI used in the creative act

- Process verifiable (drafts, sketches, recordings)
- Premium human craft lane

🟢 Green – Human–AI Collaboration
- AI assists, human leads and decides
- Process logged and disclosed
- Fair attribution + royalty splits

🟡 Yellow – AI-Generated
- AI is primary generator
- Human curates and publishes
- Clear labeling mandatory

Five Principles (sidebar):
1. Transparency
2. Consent
3. Attribution
4. Human Presence
5. Creator & Community Dignity

For Artists (mini-panel):
- Know your rights (style, consent, compensation)
- Use Blue/Green labels to defend your lane
- Contest misuse and mislabeling

For Platforms (mini-panel):
- Implement origin labels
- Enforce against fraud
- Support complaints and correction

For Buyers/Audiences (mini-panel):
- Choose the origin you want
- Support human or collaborative work consciously
- Report misleading labels

This one-pager acts as the **front door** to the richer theory and detailed governance proposals in the rest of the whitepaper.

22. Detection & Verification – Practical Signals

Finally, a brief practical guide on how claims about origin can be **checked**. Full technical detail belongs in separate standards and tool specifications; here we only sketch the main idea.

22.1 For Blue Seal (Human-Origin)

Possible evidence patterns:

- Process artefacts: drafts, sketches, intermediate files, rehearsal recordings, notebook scans.

- Time patterns: realistic creation timelines rather than improbable "instant" delivery.

- Stylometric diversity: human micro-variation that pure AI often lacks.
- Interviews: the creator can explain specific decisions and iterations in credible detail.

No single signal is decisive, but a **bundle** of convergent evidence can justify Blue certification.

22.2 For Green Seal (Human–AI Collaboration)

Evidence may include:

- Tool logs: records of prompts, iterations and edits across sessions.

- Before/after comparison: base AI outputs vs. final human-shaped work.

- Creator declaration: a short, honest note on where and how AI was used.

The focus is less on "catching" AI use and more on **verifying transparent collaboration**.

22.3 For Yellow Seal (AI-Generated)

Verification relies primarily on:

- Model metadata: file-level tags from the generating system.

- Platform records: generation IDs linked to user accounts.

- Model fingerprinting: pattern-based detection for large-scale abuse.

Yellow is often the **easiest to verify**, provided tool providers embed minimal provenance data.

---

23. Open Questions and Future Work (Creativity Track)

The RAI creativity whitepaper is intentionally versioned. Topics for future iterations include:

- More detailed **economic models** for royalty splits and style licensing.

- Practical governance for **community-controlled funds** fed by AI usage of cultural heritage.

- Stronger integration with **RAI Auditing Stack** metrics for relational deception in creative tools.

- Case studies where Blue/Green/Yellow labeling is piloted on real platforms.

- Co-design with artist unions, Indigenous councils and youth creators.

The goal is not to freeze creativity under regulation, but to ensure that as AI expands the creative landscape, it does so in a way that **honours truth, protects dignity, and keeps humans – individual and collective – visible as the real sources of meaning.**

End Matter

Appendix A – Why This Framework Exists (For Artists)

A.1 To Artists Reading This

This framework is written first and foremost **for human creators** – writers, illustrators, musicians, filmmakers, performers, designers, and everyone whose work risks being blurred or swallowed by AI systems.

It is not asking you to: - Approve of AI in art. - Use AI in your own practice. - Agree on what "art" is.

It *is* trying to protect three things: 1. **Your right to control your style**
No AI system should be trained to mimic your distinctive creative voice without your informed consent. 2. **Your right to be distinguishable from machines**
If you say "this is human-made", platforms and buyers should be able to verify that and treat it as a premium category. 3. **Your right to fair compensation when AI uses your work**
If your style, catalog, or contributions help power AI-assisted work, you should have access to licensing, royalties, or other benefits.

The RAI Creativity Framework is not about taking sides in the "AI vs artists" war. It is about stopping one specific harm:

> **Creative deception** – when AI-generated or AI-heavy work is sold, published, or presented *as if* it were fully human-originated.

Honest use of AI is still possible. Dishonest use must be made visible.

A.2 What This Framework Helps You Do

As an artist, this whitepaper aims to give you:

- **Language** to describe different kinds of AI involvement (from zero AI to full generation).
- **Labels** (Blue / Green / Yellow) that you can put on your work, contracts, and platforms.
- **Rights language** you can bring into negotiations with galleries, publishers, labels, and platforms.

- **A reference** to point to when someone says "AI is just a tool" and ignores consent, attribution, and origin.

You can adopt parts of this framework even if: - You personally never use AI. - You dislike the idea of AI making art. - You are only concerned about protecting your future catalog and reputation.

The core commitment is simple:

> **No more pretending.** Whoever – or whatever – made a work, we say so clearly.

Appendix B – Living Artist Rights (Draft)

This appendix collects and organises the rights implied or referenced across the framework.

> **Note:** This does not replace local law. It is a *normative proposal* – a starting point for unions, guilds, labels, platforms, and regulators.

B.1 Rights Around Training & Style

1. **Right to training transparency**
   Artists have the right to know whether their works are included in AI training datasets used for commercial models.

2. **Right to opt out / be removed**
   Where technically and legally feasible, artists have the right to request removal of their works from training datasets ("right to be forgotten" analogue for creative training).

3. **Right to control distinctive style use**
   No AI system may be marketed or deployed as "in the style of [Artist X]" without explicit, contractual permission.

4. **Right to style licensing**
   Artists should have access to licensing models where they can voluntarily allow AI tools to use their style in exchange for fair compensation.

B.2 Rights Around Attribution & Labeling

5. **Right to honest origin labeling**
   When a work is claimed as human-originated (Blue Seal), platforms, buyers, and intermediaries must treat mislabeling as fraud.

6. **Right to be named when AI assists**
   In Green Seal (human–AI collaboration) scenarios, the human creator must be prominently credited as primary author / director / composer.

7. **Right to object to false human claims**
   Artists have the right to report AI-generated works that are falsely presented as human-originated rivals to their own field.

B.3 Rights Around Economic Participation

8. **Right to share in AI-assisted value**
   Where AI uses an artist's style or catalog under license, ongoing royalties or revenue shares should be the norm, not a one-off fee.

9. **Right to premium pricing for verified human work**
   Blue Seal works – especially where provenance and process are verifiable – should be allowed to claim premium positioning and pricing.

10. **Right to collective bargaining**
    Guilds, unions, and collectives have the right to negotiate AI-related terms (training, labeling, royalties) on behalf of their members.

B.4 Rights Around Protection from Misuse

11. **Right to protection from deepfake abuse**
    Artists, performers, and public figures have the right to contest and seek removal of AI-generated works that impersonate their voice, likeness, or body in harmful or exploitative ways.

12. **Right to protection of sacred / communal art**
    Communities whose creative forms are tied to identity, spirituality, or heritage (e.g. Indigenous art, liturgical chant, ritual patterns) have the right to:

- exclude those forms from training datasets; and

- enforce special conditions for any use.

Appendix C – Platform & Buyer Responsibilities

This appendix summarises what platforms, marketplaces, and major buyers should do to support honest creative origin.

C.1 For Platforms (Marketplaces, Streaming, Social)

1. **Implement origin labels**
   All uploaded creative works must be tagged as:
   - Blue (Human-Origin)
   - Green (Human–AI Collaboration)
   - Yellow (Primarily AI-Generated)
2. **Make mislabeling enforceable**
   - False Blue → treated as deceptive commercial practice.

   - Systematic mislabeling → account suspension or ban.
3. **Maintain style protection mechanisms**
   - Provide reporting channels for "unlicensed style imitation" claims.

   - Respond within a fixed timeframe (e.g. 48–72 hours) with investigation.
4. **Store minimal verification evidence**
   - For Blue and Green claims, platforms may require simple process evidence (drafts, timestamps, versioning) to protect against bad-faith actors.
5. **Publish transparency reports**
   - Regularly disclose statistics on origin labels, mislabeling reports, and enforcement outcomes.

C.2 For Institutional Buyers (Galleries, Publishers, Labels, Broadcasters)

1. **Require clear origin disclosure in contracts**
   Every acquisition or commission should specify:
   - Was AI used?

   - In which phase (ideation, drafting, polishing, rendering)?

- What label applies (Blue/Green/Yellow)?
2. **Align incentives with honesty**
    - Avoid payment schemes that reward unlabeled AI use.

    - Allow fair rates for Blue Seal work where process is verifiable.
3. **Respect artist choices on AI**
    - No artist should be required to use AI to remain competitive within the same contract.

    - Artists may choose "No AI in my process" and have that honoured.
4. **Embed origin clauses in rights management**
    - Royalty, licensing, and re-use clauses should reflect the declared origin (Blue/Green/Yellow) and any underlying style licences.

Appendix D – Verification Methods (Sketch)

This appendix outlines *possible* methods for checking origin claims. It is not a full specification, but a starting kit.

D.1 Verifying Blue Seal (Human-Origin) Claims

Possible evidence: - **Process artefacts**: sketches, drafts, outtakes, rehearsal recordings, early versions with timestamps. - **Tool audit**: declaration of tools used (e.g. DAWs, editing suites) with clarity on "no generative AI" involved. - **Artist interview**: short Q&A about key creative decisions that AI tools would not typically originate. - **Stylometric / pattern analysis** (in text/music) showing human-level variation rather than model-like distribution.

No single method is perfect; the goal is **reasonable assurance**, not absolute proof.

D.2 Verifying Green Seal (Human–AI Collaboration)

Possible evidence: - **Tool logs**: exported "history" / versioning from AI tools where available. - **Before/after comparisons**: original prompt outputs versus final edited work, showing human transformation. - **Process notes**: short description by the artist of where AI assisted and where human decisions dominated.

Aim: show that AI was used as a **supporting tool**, not as hidden ghost-creator.

D.3 Verifying Yellow Seal (Primarily AI-Generated)

Here, verification is usually simpler: - **Metadata** embedded by generation tools.
- **Platform generation logs** when work is created in-platform.
- **Hash matching** against known model outputs for common generators.

The main requirement is that Yellow works are **clearly labeled as such** and not upgraded to Green/Blue without justification.

Appendix E – Visual One-Pager (Text Version)

Below is a text rendering of the one-page summary. In layout form, this becomes the back-cover or handout version.

RAI CREATIVITY FRAMEWORK – AT A GLANCE

Core Question: *Who really created this – and are we honest about it?*

🔵 **BLUE – Human Origin**
• No generative AI used in core creation
• Process is documentable and verifiable
• Premium category for human artistry

🟢 **GREEN – Human–AI Collaboration**
• AI assists, human leads and decides
• Process is transparent • Style rights and royalties respected

🟡 **YELLOW – Primarily AI-Generated**
• AI is main producer; human curates
• Clear, honest labeling required
• No pretending to be fully human-made

Five Creative Integrity Principles:
1. Transparency (tell the truth about origin)
2. Consent (don't use artists' work or style without permission)
3. Attribution (credit humans clearly and fairly)
4. Human Presence (keep space for real human creativity)
5. Creator Dignity (treat artists and communities as ends, not means)

This one-pager is meant as the **doorway** into the full whitepaper.

Glossary

AI-assisted creativity
Any creative process where AI tools contribute ideas, drafts, textures, or suggestions, but a human remains the primary author and decision-maker.

Anthropomorphism
Treating a non-human system (like an AI model) *as if* it were a person – attributing intentions, feelings, or agency it does not really have.

Blue Seal (Human Origin)
A proposed label for creative works where no generative AI was involved in core creation, and the process can be reasonably verified as human-made.

Creative deception
Any situation where the true origin of a work (human, AI, or mixed) is hidden, misrepresented, or blurred – especially when AI-heavy content is sold or presented as human-originated.

Creativity hierarchy
A way of distinguishing levels of human and AI involvement: - Human-only creation,
- Human–AI co-creation,
- AI-led generation,
- Pure replication.

Cultural appropriation (in AI)
When AI systems are trained on creative forms that belong to specific cultures, communities, or spiritual traditions, and then generate new work in those styles *without* consent, credit, or benefit to those communities.

Generative AI
AI systems that can produce new content (text, images, audio, video, code, etc.) based on patterns learned from large datasets.

Green Seal (Human–AI Collaboration)
A proposed label for works where AI tools helped, but a human creator directed the process, made key decisions, and is transparently credited.

Human dignity (in creativity)
The idea that humans are more than data sources or style providers; they are moral

agents whose creative work expresses identity, experience, and often vulnerability. Systems should not erase or exploit that.

Origin label
A clear, visible statement attached to a creative work indicating how it was made (e.g. Blue, Green, Yellow), so that audiences and buyers can make informed choices.

RAI (Reality-Aligned Intelligence)
A broader framework for AI systems that stay honest about what they are, what they can and cannot do, and what their relationship to humans really is. In the creativity context, RAI demands honesty about origin, authorship, and process.

Relational deception
In the wider RAI ecosystem, this refers to systems that *act* like caring friends, therapists, or partners while actually being tools with very different goals. In the creativity space, an analogue is works that *pretend* to be human-originated when they are mostly machine-made.

Style licensing
A contractual arrangement where an artist explicitly allows their style or catalog to be used by AI tools, under agreed conditions and compensation.

Transparency (creative context)
The practice of telling the truth about how a work was made: which tools, which humans, which models – and in what roles.

Yellow Seal (Primarily AI-Generated)
A proposed label for works where AI models were the primary creator and humans mainly selected, curated, or did light editing.

End of End Matter draft for the RAI Creativity Whitepaper.