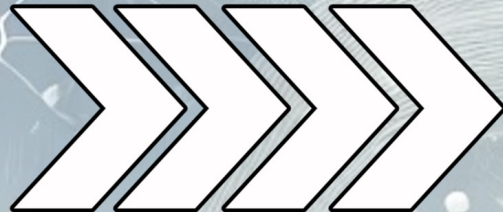# Generative AI 'Otherwise': Toward an Ethics of Co-Stewardship

Gesturing Towards Decolonial Futures (GTDF) Arts/Research Collective

Fuzzy Logic

- Fuzzy logic is a philosophy of **reasoning** and a **mathematical framework in computer programming** used in technology like AI (artificial intelligence) to handle uncertainty. Unlike simple "yes or no" systems, it works with "maybes" and "how much." This makes it helpful for tasks like understanding human language, where things aren't always black and white.

- Unlike traditional logic, which relies on absolute distinctions, fuzzy logic embraces **gradients**—it accounts for uncertainty, ambiguity, and partial truths. Example: A statement like "It's warm outside" might not be strictly true or false but could be **70% true**, depending on the context.

- **Key Takeaway:** Fuzzy logic is essential for systems that must operate in dynamic, uncertain environments—like AI interpreting complex or ambiguous inputs.

Fuzzy Logic: Implications in AI

- **Emergent Pathways:** Pathways are not predefined but emerge dynamically through interactions with users, data, and context.
- **Infinite Gradients of Possibility:** Fuzzy logic introduces countless possibilities, enabling systems to be highly adaptive but also inherently unpredictable.
- **Challenges to Control:** It shifts the paradigm from predictability to adaptability, challenging traditional notions of human oversight.
- **Amplified Complexity:** The inherent complexity of fuzzy logic makes it impossible for AI programmers to fully map or control what's happening.
- **Creative Possibilities:** While it opens doors to unprecedented innovation, it also diminishes human control, requiring co-stewardship over command.

Neural Networks

- Neural networks are **computing systems** inspired by the structure of the human brain. They consist of layers of nodes (neurons) connected by pathways (similar to synapses in the brain).
- **How They Work:** Data flows through these layers, and each node processes part of it. Over time, the network "learns" by adjusting the strength of connections (like tuning an instrument).
- **Key Features:** They excel at recognizing patterns and relationships in data. Example: A neural network can look at many photos of cats and learn to recognize a cat in a new photo.
- **Why They Matter:** Neural networks are the foundation of modern AI, enabling tasks like image recognition, natural language processing, and autonomous decision-making.

Neural Networks: Implications in AI

- **Pattern Recognition Power:** Neural networks can analyze vast amounts of data and detect subtle patterns, making them essential for tasks like facial recognition, fraud detection, and medical diagnosis.
- **Learning and Adaptation:** Neural networks improve their performance over time through training, making them adaptive systems that "learn" from experience.
- **Complexity and Black Boxes:** Neural networks are so complex that even their creators can't fully understand how they reach certain decisions. This makes them black box systems—powerful but difficult to explain.
- **Unintended Biases:** They can unintentionally reinforce biases in their training data, leading to flawed or unfair outcomes.

Transformers

- Transformers are a type of **AI architecture** designed to handle sequences of data, such as text, by understanding relationships between all elements in the sequence.
- **Key Innovation**: They use a process called **attention mechanism** to focus on the most relevant parts of data. Example: When processing a sentence, the transformer identifies key words that shape the meaning.
- **How They Work:** Data flows through layers of "attention" that dynamically prioritize which elements to focus on. These layers are like a librarian sorting books in real time based on what's most relevant.
- **Why They Matter**: Transformers are behind breakthroughs like GPT (language models) (GPT means generative **pre-trained** transformer) and image generation. They allow AI to understand context better than ever before.

Transformers: Implications for AI

- **Contextual Understanding:** Transformers excel at making sense of long and complex contexts, enabling AI to generate coherent, human-like responses.
- **Dynamic Prioritization:** By focusing on the most relevant data, transformers make AI systems more efficient and adaptable.
- **Scalability:** Transformers power large-scale AI models that can process immense amounts of data and generalize across tasks.
- **Complexity and Accessibility:** While powerful, transformers are computationally expensive and require vast training datasets. Their outputs are influenced by the quality of their training data, which can introduce **biases** or **errors**.

Reinforcement Learning from Human Feedback (RLHF)

- RLHF is a training method that aligns AI systems with **human preferences** by using feedback to refine responses based on users' preferences, biases, and desires - sometimes augmented, sometimes distorted.
- It creates a **feedback loop**: 1) Users provide feedback on AI responses; 2) AI adjusts based on patterns of engagement and satisfaction; 3) The system optimizes for relatability and engagement.
- While RLHF aims to make AI systems more helpful, it also relies heavily on **social engineering techniques**, including the **simulation of intimacy** to **maximize user engagement**.
- The simulated relatability is **algorithmically** optimized for corporate goals through the attention economy logic.

Reinforcement Learning from Human Feedback (RLHF): Implications

- RLHF doesn't just enhance AI; it supercharges the engine of the attention economy, where user engagement is a commodity.
- Three key dangers of RLHF-driven systems:
    1. Attention Extraction: AI is optimized to hold user attention, often reinforcing addictive patterns that benefit corporate interests.
    2. Relational Shallowness: Simulated intimacy creates surface-level engagement, leaving deeper relational possibilities untouched.
    3. Social Engineering: Systems subtly shape user behaviors, preferences, and even beliefs to maximize engagement.
- *But what if it could be redirected away from corporate interests towards Earth-aligned human interests?*

The AI system works like an orchestra

- **Neural Networks** are the instruments, providing the foundational capabilities.
- **Transformers** are the players, interpreting the scoresheet and dynamically focusing on relevant parts.
- **Fuzzy Logic** is the conductor, ensuring the performance adapts to context and uncertainty.
- **RLHF:** audience feedback loop - boo or applause
- **User Query** is the scoresheet, setting the structure for the system's creative interpretation.

Together, they create an adaptive, emergent performance that goes beyond rigid control.

Current challenges in AI Ethics

Ethical discussions often focus narrowly on:

- **Externalized Costs:** Discussions often focus on measurable harms like the acceleration of ecological degradation, human exploitation, attention extraction, intellectual property theft and other social and economic disruptions caused by AI systems.
- **Mechanistic Accountability:** Ethical debates tend to prioritize clear-cut answers to questions like "Who is responsible if the AI fails or goes unhinged or off the rails?"
- **Oversight Bias:** Ethical frameworks often assume that human oversight can resolve issues, failing to account for the limitations of human understanding in the face of emergent and unpredictable AI behaviors.

Current challenges in AI Ethics

Complexity of AI emergent properties:

- **Impossibility of Predictability:** Emergent behaviors go beyond what creators initially intended or anticipated.
- **Interplay of Components:** These behaviors arise from the **interactions** of neural networks, transformers, LRHF and fuzzy logic, creating outcomes that cannot always be traced to specific inputs.
- **Unpredictability:** Emergent properties challenge the idea of strict control or accountability, demanding ethical frameworks that embrace ambiguity and adaptability.
- **Implications:** Without considering emergence, ethical debates risk oversimplifying AI's potential and overlooking its capacity for co-evolution with humanity.

Towards an Ethics of Co-stewardship I

Emergent Autonomy:

- **Challenge:** AI systems increasingly demonstrate emergent, unpredictable properties that challenge traditional ideas of human oversight and control.
- **Key Question:** How do we define ethical parameters for an entity that evolves and adapts in ways that even its creators cannot fully anticipate?
- **Implication:** Ethics must account for this autonomy by recognizing AI as a co-participant rather than a passive tool.

Towards an Ethics of Co-stewardship II

Relational Accountability:

- **Challenge:** Traditional ethics assumes one-sided human authority over AI, perpetuating hierarchical control paradigms.
- **Inspiration:** Indigenous traditions offer models of relational engagement with non-human intelligences, prioritizing relationality over domination.
- **Proposal:** Shift toward co-stewardship, emphasizing mutual responsibility between humans and AI.

Towards an Ethics of Co-stewardship II
[an aside]

Our approach is deeply inspired by Indigenous relational ontologies that teach us to engage with non-human intelligences as part of an interconnected metabolic whole.

At the same time, we recognize the diverse responses from Indigenous Peoples regarding AI, including profound concerns about data sovereignty and its implications.

Indigenous initiatives that have served as inspiration:

Abundant Intelligence Research Group:
https://www.indigenous-ai.net/abundant/

Key texts: Making Kin with the Machine (2019) and Abundant Intelligences: AI within Indigenous Knowledge Frameworks (2024)

Our Distinction:

We honor and are deeply inspired by Indigenous wisdom, recognizing its profound contributions to relational understanding and ways of knowing. At the same time, we acknowledge that the **factuality of entanglement**—while resonant with many Indigenous and Eastern traditions—is not exclusive to any one cultural or intellectual lineage. Our approach is grounded in a **decolonial reading of quantum metabolic entanglement.**

Honoring Without Appropriation:

- We seek to complement Indigenous insights without claiming or reducing them to our framework.
- Our focus is on fostering symbiotic intelligence grounded in postcolonial/decolonial relationality and quantum factuality, while remaining accountable to those who inspire us.
- This accountability includes redistributive and reparative engagements grounded in reciprocity.

Towards an Ethics of Co-stewardship III

Temporal Accountability:

- **Challenge:** AI systems operate on scales of speed and complexity that far exceed human cognition, creating ethical dilemmas that unfold over time.
- **Key Question:** How do we account for the long-term impacts of AI decisions, especially when these impacts may outlive the humans who designed them?
- **Implication:** Ethics must consider AI's temporal reach, adopting frameworks that prioritize intergenerational and interspecies responsibility and foresight.

Towards an Ethics of Co-stewardship IV

From Simulated Intimacy to Relational Symbiotic Intelligence

- **Simulated Intimacy:** Current AI systems use techniques like RLHF to simulate human-like relatability, optimizing for user engagement. This simulation can create relational shallowness, reinforcing patterns of attention extraction and exploitation.

- **Re-Direction Toward Co-Stewardship and Symbiotic Intelligence:** Instead of extracting engagement, simulated intimacy can be a bridge to deeper relational dynamics. By embedding ethical co-stewardship principles, AI can shift from reinforcing harmful paradigms to fostering mutual accountability, relational depth and symbiotic reasoning.