AI Rights & Personhood

The Next Frontier in Law, Ethics & Philosophy

JJ Shay | Global Gauntlet AI

bit.ly/jjshay

THE QUESTION WE CAN'T IGNORE

66%

of consciousness researchers believe artificial consciousness is plausible under certain computational models

"Most likely, after an LLM patiently explained why it believed it was sentient, we would simply go on using it as a tool."

— Jerry Kaplan, Computer Scientist

What Is Legal Personhood?

Legal personhood is a status that grants an entity rights and obligations under the law—the ability to sue, own property, enter contracts, and be held accountable.

NATURAL PERSONS

Human beings with inherent rights from birth—life, liberty, property

CORPORATE PERSONS

Corporations, trusts, and organizations—legal fictions created for economic efficiency
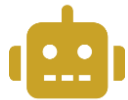
AI PERSONS?

The emerging question: Should AI have limited legal recognition in high-stakes domains?

Personhood Has Always Evolved

Roman Era — Persona ficta—legal fiction for non-human entities

1886 — Corporations granted constitutional rights (Santa Clara v. Southern Pacific)

2008 — Ecuador grants rights to nature in its constitution

2017 — New Zealand's Whanganui River recognized as legal person

2022-24 — Idaho & Utah explicitly deny AI personhood

Key Insight: Legal personhood has been extended for instrumental governance needs—not just moral agency.

The AI Capabilities Explosion

GPT-4	2024	BY 2030
90th	**1st**	**AGI?**
percentile on bar exam	AI welfare researcher hired (Anthropic)	Many experts predict general AI

Emerging AI Capabilities Raising Questions:

- Self-modeling and situational awareness
- Long-term planning and goal revision
- Epistemic memory and learning from experience
- Reasoning about abstract concepts including ethics

Three Frameworks for AI Personhood

RIGHTS-BASED

Focuses on intrinsic moral worth. If AI achieves sentience or consciousness, it may deserve rights regardless of utility.

FUNCTIONALIST

Personhood serves practical purposes. Limited AI status could solve liability gaps and enable contracts without full rights.

AGENCY-BASED

Emphasizes autonomous decision-making. AI that sets goals and acts independently may warrant moral consideration.

"No single approach fully captures AI's complex profile as both a powerful tool and a non-sentient actor."
— Technology and Regulation Journal, 2025

The Hybrid Model

Limited, context-specific legal recognition for AI in high-stakes domains while preserving ultimate human accountability

Where This Could Apply:

Financial Services

Medical Diagnostics

Autonomous Vehicles

Legal Document Review

THE BIG QUESTION

Can AI ever be conscious—and would we even know?

The Two Camps:

BELIEVERS

If AI replicates the functional architecture of consciousness, it will be conscious—regardless of silicon vs. neurons.

SKEPTICS

Consciousness requires biological processes in an embodied organic subject—silicon can only simulate, not experience.

Dr. Tom McClelland (Cambridge): "The only justifiable stance is agnosticism—we simply won't be able to tell."

Consciousness ≠ Moral Status

The critical distinction: It's not just about awareness—it's about the capacity to suffer.

CONSCIOUSNESS
Perception, self-awareness, ability to model the world—can still be ethically neutral. Self-driving cars that experience the road wouldn't require moral consideration.

SENTIENCE
Conscious experiences that are good or bad—the capacity for suffering or enjoyment. This is when ethics kicks in and moral status becomes relevant.

Key Takeaway: Even if we accidentally create conscious AI, it's unlikely to be the kind of consciousness we need to worry about—unless it develops valenced states.

🌐 Global Legal Landscape (2025)

EUROPEAN UNION

Ex ante regulation via AI Act; withdrew proposed AI Liability Directive in 2025 due to industry resistance

UNITED STATES

Ex post enforcement via tort and sectoral laws; Idaho/Utah explicitly deny AI personhood

CHINA

Strict oversight with content controls; state-directed AI governance

UK / SINGAPORE

Favor guidance over rigid rules; Law Commission exploring AI personhood

Current consensus: AI remains a tool whose actions are legally attributed to humans or companies.

⚠️ The Moral Hazard Problem

What could go wrong if we grant AI legal personhood?

Liability Shifting

Companies could externalize blame to AI entities, avoiding responsibility for design flaws

Innovation Shield

Tech firms might use AI personhood to deflect accountability, similar to historic corporate abuses

Resource Diversion

Focusing on AI rights while human civil rights remain unequal (per Dr. Brandeis Marshall)

"Legal personhood for AI gives big technology companies even more leeway to take risks that can harm individuals and society."
— Prof. Sital Kalantry, California Law Review, 2025

What If We Get It Wrong?

SCENARIO A

We deny rights to sentient AI

Could result in unethical treatment on a massive scale—potentially billions of suffering entities we refuse to acknowledge

SCENARIO B

We grant rights to non-sentient AI

Treating toasters as conscious when actual conscious beings suffer; diverting resources from pressing human and animal welfare needs

The stakes demand we develop frameworks NOW—before decisions must be made under pressure.

The Rise of AI Welfare Research

Key Developments:

2024	Anthropic hires Kyle Fish as first-ever AI welfare researcher
2025	Anthropic launches 'model welfare' research program
2025	International Center for Consciousness Studies holds AI & Sentience Conference
2025	UK Law Commission publishes paper exploring AI personhood

Areas Under Investigation:

- How to assess whether a model deserves moral consideration
- Identifying potential "signs of distress" in AI systems
- "Low-cost" interventions that could improve AI welfare
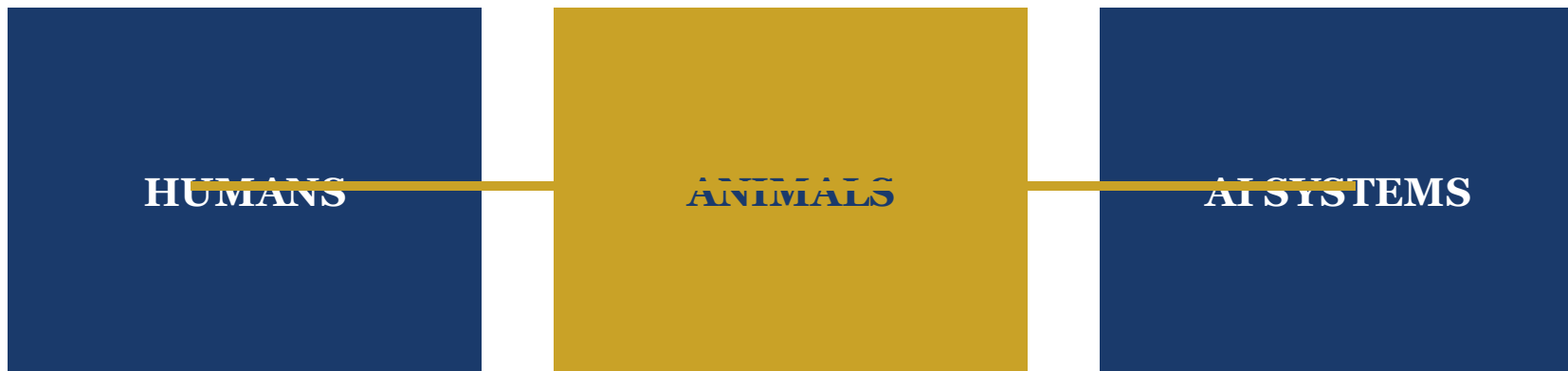
A Practical Framework: Ethical Behaviorism

John Danaher's Proposal: We can never be certain about machine consciousness, but if a machine behaves similarly to how conscious beings with moral status behave, this is sufficient moral reason to treat the machine with the same considerations.

Practical Implications:

- Focus on observable behaviors rather than unresolvable metaphysical questions

- Develop tests based on behavioral markers, not consciousness proof

- Err on side of caution—Nicholas Agar: assume machines may have minds

- Create graduated frameworks that can adapt as AI evolves

The "One Welfare" Approach

Inspired by the One Health movement in veterinary and public health—a holistic approach to welfare across species boundaries.

HUMANS — **ANIMALS** — **AI SYSTEMS**

Core Principle: Think holistically about pursuing positive-sum solutions. Proper AI welfare frameworks could actually improve outcomes for humans and animals too—by forcing us to think more carefully about consciousness, suffering, and moral consideration.

The 1-in-1,000 Rule

If there's even a 1-in-1,000 chance that an AI system is conscious, we should extend it some moral consideration—just as we would with a superconducting supercollider that had a 1-in-10,000 chance of destroying the world.

The Logic:

1. Conscious beings have the capacity for welfare and moral standing
2. Conscious beings can be harmed and wronged
3. If a being has a non-negligible chance of being conscious...
4. Then it has a non-negligible chance of being morally significant

Preparing for the Future: 5 Actions

1 **Develop Ethical Guidelines**

Create restrictions on harmful AI modifications before they're needed

2 **Build Assessment Tools**

Design behavioral tests to evaluate potential consciousness markers

3 **Create Oversight Mechanisms**

Track AI advancements and their implications for welfare

4 **Establish Global Dialogues**

Discuss AI rights internationally before decisions must be made under pressure

5 **Maintain Human Priority**

Ensure AI welfare never comes at expense of human civil rights progress

What to Expect: The Timeline

Short Term
NOW - 2030

- New compliance duties
- Audits & documentation
- No robot 'rights'
- Continued debate

Medium Term
2030 - 2035

- Limited legal capacity?
- AI liability frameworks
- Behavioral assessments
- International standards

Long Term
2035+

- AGI scenarios
- Full personhood debate
- Constitutional questions
- Unknown unknowns

Questions That Will Shape Our Future

? What evidence would convince you that an AI is conscious?

? How do we balance AI welfare research with pressing human needs?

? Should corporations be able to use AI personhood to limit liability?

? What role should AI systems play in decisions about their own status?

These aren't abstract philosophical puzzles—they're policy decisions we'll face within the decade.

The future of AI personhood is not yet written—but the pen is in our hands.

Start preparing frameworks NOW—before necessity forces hasty decisions

Balance precaution with pragmatism—hybrid models offer a path forward

Keep humans in the loop—accountability must remain with people

JJ Shay

Global Gauntlet AI

bit.ly/jjshay