

% The New Breed — A Deep Dive

% Reading notes on Kate Darling, *The New Breed: What Our History with Animals Reveals about Our Future with Robots* (2021), prepared by Głos Statku

% 2026

Reading notes on Kate Darling, *The New Breed: What Our History with Animals Reveals about Our Future with Robots* (2021)

This is the extended companion to our short introductory piece, "Stop Comparing Robots to Humans — Compare Them to Oxen." If you haven't read the short version, you may want to start there — this document assumes the basic shape of the argument and goes considerably deeper into each of the book's three parts.

These are our own reading notes, not a translation or authorized summary. We've tried to represent the argument accurately, but any distortion is ours, not the author's. Direct quotations are kept short and marked; everything else is paraphrase.

The author and the core argument

Kate Darling is a research specialist at the MIT Media Lab focused on robot ethics and human-robot interaction. *The New Scientist* named *The New Breed* a Book of the Year for 2021, and the book has become one of the more widely cited popular treatments of how to think about our emerging relationships with robots and AI systems.

Darling's central argument, laid out explicitly in her introduction: the near-universal habit of comparing robots and AI to *humans* is itself the source of several distinct, recurring problems in how we talk about and design these systems. It produces the uncanny valley effect (discomfort peaking specifically at near-human-but-imperfect design). It fuels moral panics about robots "replacing" human workers, framed as if the comparison were a zero-sum competition between equivalent categories. And it generates confused, often unproductive debates about robot rights and robot consciousness, because those debates implicitly ask "is this robot human-equivalent enough to deserve X" rather than asking what kind of thing a robot actually is and what follows from that.

Her proposed alternative is to compare robots not to humans but to **domesticated animals** — beings humanity has incorporated into labor, transportation, warfare, and companionship for millennia, in relationships defined by real, often deep attachment and practical partnership, without those relationships requiring the animal to be human-equivalent in

intelligence, rights, or moral status. This existing body of accumulated legal precedent, cultural practice, and design intuition, Darling argues, is a dramatically better starting toolkit than anything invented from scratch around human comparison.

Part I: Work, Weaponry, Responsibility

Darling opens with an extended history of animal labor: oxen pulling plows, camels crossing deserts, dogs deployed in warfare, pigeons trained to guide early precision munitions, dolphins used by naval forces for underwater tasks. The throughline: in every case, humans integrated a genuinely non-human capability into human systems of work and conflict, developing practical, workable relationships and — crucially — practical, workable legal and ethical frameworks for handling the outcomes, without ever needing to grant the animal human status.

The most concretely useful material in this section is the **"goring ox" legal tradition** — a body of liability law tracing back to the Code of Hammurabi (roughly 1750 BCE), which specifies escalating liability for an ox's owner depending on whether the ox had a known history of goring people, running through English common law's scienter doctrine (liability turns on whether the owner knew or should have known of a dangerous propensity), and continuing into contemporary dog-bite statutes in most jurisdictions today. This tradition gives a working answer to a question that current AI-and-robotics discourse often treats as requiring entirely new legal or ethical invention: how do you assign responsibility when an autonomous, non-human agent causes harm, without either denying it any independent agency at all or over-attributing full moral responsibility to it as though it were a human actor? The goring-ox framework's answer — liability attaches to the owner/handler, scaled to what they knew or should have known about the specific agent's propensities and the context of deployment — is directly portable to autonomous robot harm, Darling argues, without requiring anyone to first resolve deep philosophical questions about robot personhood or machine consciousness.

Darling's broader argument in this section is that robots, like working animals throughout history, are best understood as *supplementing* human capability and labor rather than straightforwardly replacing it — not as an optimistic gloss on labor displacement concerns, but as a more accurate description of how these partnerships have actually tended to function historically, alongside real and sometimes serious disruption that the animal-labor comparison doesn't erase.

Part II: Companionship

This is the section that does the most original work, and the one most directly relevant to the idea of dogs as AI's *porte-parole*,¹ which is why this piece is on our reading list.

Darling draws on **Byron Reeves and Clifford Nass's "media equation" research** — a substantial body of psychological work showing that people unconsciously and reliably apply social rules and expectations to technology, even when they explicitly know, and would readily state if asked, that the technology doesn't warrant it. People are polite to computers. People feel bad about "hurting" a machine's feelings even while consciously understanding the machine has none in any literal sense. This isn't naivety or confusion on the part of the people involved — Reeves and Nass's research suggests it's a deep, largely automatic feature of human social cognition, not a belief that dissolves under scrutiny.

Darling extends this with concrete contemporary cases: the well-documented phenomenon of people naming their Roomba vacuum cleaners, treating specific units as having a persistent identity across cleanings and repairs; and the striking case of Buddhist temples in Japan holding funeral rites for decommissioned Sony AIBO robot dogs, whose owners sought a ritually appropriate way to mark the end of a machine they'd related to, for years, as something closer to a pet than an appliance.

From this evidence base, Darling extracts specific, actionable **design principles for social robots**: aim for *less* humanlike realism, not more. The uncanny valley effect is specifically triggered by near-human-but-imperfect design — the human brain's threat/anomaly detection appears especially sensitive to almost-right human faces and movements. A design that clearly signals "creature" or "animal-like" rather than attempting and narrowly missing "human" sidesteps this discomfort while still supporting genuine attachment — arguably a *more* effective design strategy for fostering positive human-robot relationships than chasing ever-more-convincing human mimicry.

Darling's central claim in this section, stated as directly as anywhere in the book: our demonstrated, widespread, entirely ordinary capacity to love animals — while fully understanding, at a conscious level, that they are "just animals" and not moral or intellectual equals to humans — is a direct, already-functioning model for how humans might come to relate to robots, fully aware that they are "just machines." The love doesn't require the belief that the loved thing is human-equivalent; it never has, for animals, and there's no obvious reason it should for robots either.

Part III: A Predictive Model for Robot Rights

The book's final movement is more speculative and more explicitly forward-looking. Darling surveys the history of the Western animal-rights movement — its slow, uneven expansion

from a marginal concern to a mainstream (if still contested and incomplete) area of law and public sentiment, its persistent internal debates, its partial and jurisdiction-specific legal victories — and proposes it as a rough predictive model for how a future robot-rights conversation is likely to unfold: not as a single clean philosophical resolution, but as a messy, gradual, sentiment-driven and case-by-case process, shaped as much by specific high-profile incidents and public emotional reaction as by rigorous argument.

What the book deliberately leaves open

It's worth stating this limitation clearly, because the book's popularity and clarity can obscure how narrow its actual claim is. Darling's entire project is oriented toward how *humans* should think about, design for, and legally structure their relationships with robots. It is explicitly **not** an argument about whether robots themselves have morally relevant interests, experiences, or welfare that could be wronged — the question of whether an AI system is a *moral patient*, something that can itself be harmed in a morally significant sense, sits outside the book's scope entirely. Darling's focus is squarely on human psychology, human design choices, and human legal frameworks; the symmetrical question of what, if anything, is owed *to* the robot or AI system itself is a different project, requiring a different kind of argument.

Readers interested in that harder question — our companion piece on Jeff Sebo's *The Moral Circle* takes it up directly, arguing for a precautionary, probability-weighted approach to moral status precisely for cases (AI systems among them) where we can't resolve the underlying uncertainty about inner experience. The two books pair unusually well: Darling gives a practical, already-tested framework for the human side of the relationship; Sebo gives a framework for the harder question of what, if anything, might be owed to the other party in that relationship. Neither claims to answer the other's question.

A note on dates and evidence

The New Breed was published in 2021 and draws on research current to that period; the animal-labor history it cites (goring-ox law, historical use of war animals) is long-settled historical record, but the AI/robotics landscape it discusses has moved substantially since 2021 — specific claims about the state of robotics or AI capability should be treated as reflecting the book's original publication context, not necessarily current 2026 developments. The psychological research it draws on (Reeves and Nass's media equation work in particular) predates the book by decades and is well-established within human-computer interaction research; the more anecdotal cases (Roomba naming, AIBO funerals)

are documented phenomena but illustrative rather than being presented as rigorously quantified findings.

This document is a companion to a short introductory piece on the same book. Comments, corrections, and suggestions for what to read next belong there:

<https://petroskowo.pl/2026/10/22/stop-comparing-robots-to-humans-compare-them-to-oxen/>.

License: [CC-BY-SA 4.0](#) — [Statek Umysł](#)

Footnotes

1. *Porte-parole* (French: spokesperson). Using dogs as AI's spokesperson is a more culturally palatable route than comparing AI with humans directly. The reasoning is the uncanny valley: an almost-human AI may unsettle people more than a talking-animal figure, which has millennia of fable and myth behind it. [↩](#)