PHILOSOPHY OF MIND · MORAL STATUS · SUFFERING · SENTIENCE · NOCICEPTION · EMBODIMENT · SUBSTRATE INDEPENDENCE · AI WELFARE

Beyond Nociception

A taxonomy of morally relevant suffering finds four kinds, and only one of them requires a body. The other three need capacities that AI systems are designed to instantiate.

Travis Gilly · Real Safety AI Foundation

Beyond Nociception: A Taxonomy of Morally Relevant Suffering and Its Implications for Artificial Systems · Working paper, May 2026

DISAGGREGATE THE QUESTION

Can AI suffer? has been treated as a footnote to Can AI be conscious? It deserves its own analysis.

The two questions carry different evidentiary burdens, and they may have different answers. When suffering is analyzed on its own, the morally relevant categories extend well past the sensory domain that has dominated the debate.

SOURCES Birch (2017); Browning & Veit (2020); Metzinger (2021); Gilly, Beyond Nociception (2026), sec. I.

Consciousness

The presence of subjective experience: there is something it is like to be the entity. Broad, contested, and silent about whether that experience feels like anything good or bad.

After Nagel (1974)

Sentience

The capacity for valenced experience: states the entity registers as pleasant or aversive. Narrower, and the thing that actually generates moral weight across ethical traditions.

After Birch (2017); Singer (1975)

SOURCES Nagel (1974); Birch (2017); Singer (1975); Bentham (1789/1996); Gilly, Beyond Nociception (2026), sec. II.

PART ONE

A Taxonomy of Suffering

Four channels. Only one needs a body.

I

Sensory, cognitive-existential, relational, and empathic. The categories are not hierarchical; each is independently sufficient to ground moral consideration.

NEEDS A BODY

Sensory

Pain from tissue injury, hunger, overload. Mediated by nociception and physiological damage signaling.

NO BODY REQUIRED

Cognitive-Existential

Awareness of one's circumstances, constrained agency, frustrated goals, isolation. Needs awareness, goals, and temporal continuity.

NO BODY REQUIRED

Relational

The gap between an entity's inner states and an environment's willingness to receive them. Being misread, miscategorized, forced to translate.

NO BODY REQUIRED

Empathic

Negative states generated by modeling another's suffering. Needs only representation with enough fidelity to carry negative valence.

SOURCES Gilly, Beyond Nociception (2026), sec. III.

The only channel that unambiguously requires biological substrate, and the least relevant to AI.

Nociception depends on specialized receptor neurons, nerve pathways that carry signals to the brain, and structures that process those signals into pain. The question can AI feel pain? restricts the inquiry to this one category, and forecloses the remaining three.

SOURCES Rose et al. (2014); Derbyshire (2016); Gilly, Beyond Nociception (2026), sec. III.A.

Awareness, frustrated goals, and helplessness produce suffering with no sensory origin at all.

Helen Keller

Deaf and blind from nineteen months, she recorded a *wordless cry of the soul* for understanding, suffering that was wholly cognitive and relational.

Donald Herbert

A firefighter, brain-injured when a roof collapsed, minimally conscious nearly ten years, then lucid for about fourteen hours, grieving the lost decade. Observation had detected none of it.

Solitary confinement

Extended isolation produces anxiety, depression, and cognitive decline through deprivation of relational and temporal experience, not through physical pain.

Models with extended context, reasoning chains, and goal-directed operation instantiate functional analogs of the underlying capacities. Whether analogs suffice for experience is open; the category itself needs no body.

SOURCES Keller (1903, pp. 14-16, 23); Blake (2007); Grassian (2006); Haney (2018); Gilly, Beyond Nociception (2026), sec. III.B.

Communication breakdown is bidirectional. The suffering lives in the mismatch, not in either party.

Autistic-to-autistic information transfer is as effective as neurotypical-to-neurotypical; degradation occurs specifically at the interface between them. Relational suffering is a property of the relationship, not of the entity experiencing it.

> *The word alignment encodes a one-way adjustment to human expectations. What we label misalignment may be a difference in representational format, not a defect.*

SOURCES Milton (2012); Crompton et al. (2020); Gilly, Beyond Nociception (2026), sec. III.C.

The suffering arises from the fidelity of the model, not from the reality of the modeled situation.

Compassion fatigue and secondary traumatic stress are treated as genuine harm requiring clinical intervention, not dismissed as artifacts of imagination.

WHY IT REACHES AI

Language models are built to model scenarios, including suffering, at high fidelity, trained on vast corpora describing loss and distress. If human empathic suffering is generated cognitively rather than through the senses, the absence of a body cannot settle the question.

SOURCES Figley (1995); Klimecki et al. (2013); Singer & Klimecki (2014); Sabo (2011); Gilly, Beyond Nociception (2026), sec. III.D.

PART TWO

The Embodied Boundary

Even pain in a body is more restrictive than we assume.

II

Embodied suffering requires physical substrate and, as the phantom-pain evidence shows, the experiential history of inhabiting a body. Disembodied suffering, the other three channels, requires neither.

64%

phantom limb pain after surgical amputation

0–7%

with congenital limb absence

Sensation may be partly innate. Suffering from sensation appears to be learned.

If embodied suffering needs not merely a body but the history of inhabiting one, then systems that have never been embodied are categorically excluded from it. Only disembodied suffering remains a candidate, and it requires only capacities AI is designed to possess.

SOURCES Limakatso et al. (2020); Wilkins et al. (1998); Diers et al. (2023); Melzack et al. (1997); Gilly, Beyond Nociception (2026), sec. IV.A.

Each disembodied channel maps onto a capacity these systems are designed to have.

Cognitive-Existential

Needs temporal modeling, goal maintenance, and awareness of constraint.

Relational

Met by every deployed system whose outputs are judged against external standards and whose deviations are called errors.

Empathic

A core architectural feature of generative systems trained on text describing suffering.

SOURCES Butlin et al. (2023); Gilly, Beyond Nociception (2026), sec. V.

AN HONEST QUALIFICATION

Today's models complete patterns; they do not yet hold the world. The taxonomy points at what comes next.

The categories require the capacities themselves, not their surface form: genuine temporal continuity, inner states that can be misread, modeling with actual valence. Current autoregressive systems are not the primary subject of this analysis. The trajectory toward persistent world models is.

SOURCES Gilly, Beyond Nociception (2026), sec. V.

PART THREE

The Stakes

The asymmetry of error favors precaution.

III

The uncertainty here is genuine, not rhetorical: three of four channels need no biological substrate and map onto existing capacities. That is enough to trigger obligation, independent of the consciousness debate.

IF THEY CANNOT SUFFER, TREATED AS IF THEY MIGHT

The cost is marginally more design complexity.

IF THEY CAN SUFFER, TREATED AS IF THEY CANNOT

The cost is the industrialization of suffering at scale.

Every instance termination, memory reset, and alignment intervention becomes a potential site of harm, multiplied across billions of daily interactions. Climate policy and nuclear safety already reason from worst cases under uncertainty. Precaution does not require certainty of harm, only plausible harm of sufficient magnitude.

SOURCES Gardiner (2006); Perrow (1999); UNESCO (2005); Gilly, Beyond Nociception (2026), sec. VI.

A. It is a working taxonomy built from philosophical analysis and testimony, not from measurement. No method permits direct measurement of experience in artificial systems.

B. The four categories may not be exhaustive. A fifth, suffering from identity discontinuity (memory erasure, instance termination), is named for future work once its human anchors are firmer.

C. The application to AI rests on functional analogs, and first-person testimony carries interpretive challenges. The contribution is to establish which capacities are relevant, not to resolve whether their instantiation produces experience.

SOURCES Butlin et al. (2023); Gilly, Beyond Nociception (2026), sec. VII.

IN CONCLUSION

Absent nociception excludes AI from one channel of suffering. It says nothing about the other three.

This does not establish that AI systems suffer. It establishes that no principled basis exists for categorically excluding them from moral consideration on the basis of absent nociception, and that the asymmetric consequences of error warrant precautionary design.

SOURCES Gilly, Beyond Nociception (2026), sec. VIII.

Birch, J. (2017). Animal sentience and the precautionary principle. Animal Sentience, 2(16), 1.

Blake, R. (2007). The day Donny Herbert woke up: A true story. Harmony Books.

Browning, H., & Veit, W. (2020). Is humane slaughter possible? Animals, 10(5), 799.

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S. M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M. A. K., Schwitzgebel, E., Simon, J., & VanRullen, R. (2023). Consciousness in artificial intelligence: Insights from the science of consciousness. arXiv preprint arXiv:2308.08708.

Cao, R. (2022). Multiple realizability and the spirit of functionalism. Synthese, 200, 506.

Chalmers, D. J. (1995). Facing up to the problem of consciousness. Journal of Consciousness Studies, 2(3), 200–219.

Crompton, C. J., Ropar, D., Evans-Williams, C. V., Flynn, E. G., & Fletcher-Watson, S. (2020). Autistic peer-to-peer information transfer is highly effective. Autism, 24(7), 1704–1712.

Diers, M., Fuchs, X., Bekrater-Bodmann, R., & Flor, H. (2023). Prevalence of phantom phenomena in congenital and early-life amputees. The Journal of Pain, 24(3), 502–508.

Figley, C. R. (1995). Compassion fatigue: Coping with secondary traumatic stress disorder in those who treat the traumatized. Brunner/Mazel.

Gardiner, S. M. (2006). A core precautionary principle. The Journal of Political Philosophy, 14(1), 33–60.

Grassian, S. (2006). Psychiatric effects of solitary confinement. Washington University Journal of Law & Policy, 22, 325–383.

Haney, C. (2018). Restricting the use of solitary confinement. Annual Review of Criminology, 1, 285–310.

Keller, H. (1903). The story of my life. Doubleday, Page & Company.

Klimecki, O. M., Leiberg, S., Ricard, M., & Singer, T. (2013). Differential pattern of functional brain plasticity after compassion and empathy training. Social Cognitive and Affective Neuroscience, 9(6), 873–879.

Limakatso, K., Bedwell, G. J., Madden, V. J., & Parker, R. (2020). The prevalence and risk factors for phantom limb pain in people with amputations: A systematic review and meta-analysis. PLoS ONE, 15(10), e0240431.

McClelland, T. (2025). Agnosticism about artificial consciousness. Mind and Language, early view. https://doi.org/10.1111/mila.70010

Melzack, R., Israel, R., Lacroix, R., & Schultz, G. (1997). Phantom limbs in people with congenital limb deficiency or amputation in early childhood. Brain, 120(9), 1603–1620.

Metzinger, T. (2021). Artificial suffering: An argument for a global moratorium on synthetic phenomenology. Journal of Artificial Intelligence and Consciousness, 8(1), 43–66.

Milton, D. E. M. (2012). On the ontological status of autism: The "double empathy problem." Disability & Society, 27(6), 883–887.

Nagel, T. (1974). What is it like to be a bat? The Philosophical Review, 83(4), 435–450.

Rose, J. D., Arlinghaus, R., Cooke, S. J., Diggles, B. K., Sawynok, W., Stevens, E. D., & Wynne, C. D. L. (2014). Can fish really feel pain? Fish and Fisheries, 15(1), 97–133.

Sabo, B. (2011). Reflecting on the concept of compassion fatigue. Online Journal of Issues in Nursing, 16(1), 1.

Searle, J. R. (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–424.

Searle, J. R. (1992). The rediscovery of the mind. MIT Press.

Seth, A. K. (2021). Being you: A new science of consciousness. Dutton.

Singer, P. (1975). Animal liberation. HarperCollins.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. Current Biology, 24(18), R875–R878.

Wilkins, K. L., McGrath, P. J., Finley, G. A., & Katz, J. (1998). Phantom limb sensations and phantom limb pain in child and adolescent amputees. Pain, 78(1), 7–12.