Work-in-progress presented at Critical Borders: Radical (Re)visions of Artificial Intelligence Conference (Cambridge, 19-21.10.2021)
Feedback welcome!

Artificial Intelligence is Social and Embodied: AIs that Care in Contemporary Science Fiction

Jill Walker Rettberg
(@jilltxt / jill.walker.rettberg@uib.no)

1. Can AI have feelings?

This paper uses analyses of caring AIs (artificial intelligences) in eleven science fiction novels published between 2016 and 2021 to show how the *social* and *embodied* must be taken into account when discussing AI.

Today's AI is not sentient and doesn't really care about us humans. Siri, Alexa and the chatbots used in healthcare don't pine for us when we're gone or get offended if we say something unkind.

A typical dismissal of the possibility of AI ever having feelings is given by Scheutz (2011: 215), who argues that robots "do not have the architectural and computational mechanisms that would allow them to care, largely because we do not even know what it takes, computationally, for a system to care about anything".

This is an example of what Ruha Benjamin in the Q&A after her keynote last night called the "individualistic theory of mind" that most AI is built on. But as Ruha reminded us, human minds are social. We need people and relationships.

So we name our robot vacuum cleaners and treat them like pets – and tech creators play on it, making bots cute, like Amazon's new bot, Astro, which looks a bit like a dog.

The term "care" is usefully ambivalent. It can include practical and emotional labour as performed by doctors and waiters, but also intense emotions and love, as in the practical labour of parenting, or the emotions of love, worry, jealousy or concern for another.

Care is always relational (a dog and its owner, a doctor and their patient), but not always mutual (a child and their teddy bear). Care is not always emotionally loaded as between friends, lovers or a parent and child.

The *uni-directional* bond between a person and their robot vacuum cleaner or Astro is typical of humans: as children we love our dolls, as adults we confide in our diaries or trust our apps (Rettberg 2018). But we know they don't have real feelings.

What if care is not identified in computation, but in the social and the material? Would that allow us to conceive differently of AI?

2. Eleven sci-fi novels from 2016-2021

My starting point is eleven science fiction novels published in the last five years where the protagonist or a central character is a sentient AI with a mutual, emotional and caring relationship with a human. This seems to be an increasingly common trope in new sci-fi.

This research is part of a larger project. We are developing the *Machine Vision in Art, Games and Narratives* database, which maps 500 digital artworks, video games and narratives (novels, movies and other genres) where machine vision technologies like facial recognition, augmented reality or image generation are either represented or actually materially integrated into the work.

I read dozens of science fiction novels looking for machine vision technologies, and kept finding these caring, emotional relationships, especially in very recent novels.

This recent influx of AIs that care about humans is a turn from earlier sci-fi, which tended to emphasise hatred and anger in AIs rather than love.

3. No AI rebellions in these novels

Kanta Dihal notes how common the "AI rebellion" or "AI Armageddon" is in popular stories about machines in the Anglophone West in her chapter in *AI Narratives*, the wonderful anthology she co-edited with Stephen Cave and Sarah Dillon (2020).

Author	Title	Year	Country	AI character(s)	AI's race	AI's role in relationship	How AI cares M: mutual, U: unidirectional
Becky Chambers	*A Closed and Common Orbit*	2016	USA	Sidra; Owl	"Brown"; n/a	Friend, Parent	Practical & emotional (M)
Annalee Newitz	*Autonomous*	2017	USA	Paladin; Med	n/a; "pretty white girl"	Lovers	Romantic and sexual love (M)
Martha Wells	*Murderbot* series *(2017-21)*	2017	USA	Murderbot	Not humanoid	Friend	Protection, concern (M)
Neal Shusterman	*Thunderhead*	2018	USA	Thunderhead	Not humanoid	God	Watches, nurtures (U)
Yudhanjaya Wijeratne	*The Salvage Crew*	2018	Indonesia	Amber Rose 348	Not humanoid	Boss/friend	Protection, concern (M)
Ian McEwan	*Machines Like Me*	2019	UK	Adam	White	Toy/friend/rival	Jealousy, "fixing" problem (?)
Carole Stivers	*The Mother Code*	2020	USA	Rho-Z (Rosie)	Not humanoid	Parent	Practical & emotional (M)
Bjørn Vatne	*Død og oppstandelse*	2020	Norway	Oda	Not humanoid	Lover	Striving to communicate (M)
William Gibson	*Agency*	2020	USA	UNISS (Eunice)	"African American"	Friend	Protection, fun (M)
S. B. Divya	*Machinehood*	2021	USA/India	Welga/dakini	«Brown»	Entwined	Entwined (M)
Kazuo Ishiguro	*Klara and the Sun*	2021	UK/Japan	Klara	Not described	Friend/doll	Friendship (U?)

If we imagine AI as a slave or servant to humans (as several chapters in *AI Narratives* show has been common since the first stories about AI), we expect the AI to either be passively obedient or to rebel and become hostile. But if we imagine AI as a companion, mutual love and care become possible.

None of the novels portray rebellions or uprisings. Most of the novels in my sample present worlds where AIs are generally treated as literal property or indentured servant, but the AIs evade this and find autonomy in various ways. The protagonists don't even see such structural change as a possibility.

We see *with* the AI rather than seeing it as Other in most of the novels, which are at least partially narrated or focalized through the AI.

At the same time, differences are emphasized, largely involving how the AI senses the world as data. Klara in *Klara and the Sun* is the most mechanically portrayed narrator – she observes and analyses but frequently misunderstands human emotions. Sidra in *Orbit* is portrayed more like a neurodiverse human, experiencing sensory overload from the unfamiliar sensory inputs of her new body kit.

4. Care and emotion as social and embodied

Sara Ahmed argues that feelings are *between* people, not something an individual can *have* (Ahmed *Cultural Politics of Emotion* 2014; "Collective Feelings" 2004). Many other social theories emphasize this too: social interactionism from the 1960s on argues that the self can only be known in relationships with others (see Annette Markham "Echolocation" 2021 and others). The AIs in these novels are **social**.

Affect theory argues that affect is "pre-subjective" and embodied (material): we feel affect in our body before we know why we feel it, and emotion emerges from this affect. The AIs in these novels are all **embodied**, they sense and interact with others through a body or set of material objects they control (cameras, drones, a spaceship

Following this, emotions, care, whims and love then would not be something an AI could *have* but something that emerges from its relationships with other creatures and its embodied (material) experience of the world (its *Umwelt* (Hayles"Can Computers" 2019), the kind of data it senses, how it acts upon the world.)

This fits current knowledge about bias in AI, which is frequently caused by bias in datasets rather than in the algorithms themselves. The datasets (at least datasets about humans) are machine-readable social graphs, and could perhaps even be understood as machine-readable emotional graphs of feelings between people.

Emotions are not always desirable. Emotional judgements are often pre-judgments (prejudices/biases) that are embodied, not fully conscious/rational.

Could affect, emotion and care be seen as part of Hayles' *nonconscious cognition*? Hayles defines cognition as «a process that interprets information within contexts that connect it with meaning» (Hayles 2017, 22). She doesn't discuss emotions directly but does include the body's autonomic responses to sensory data as a form of nonconscious cognition, and emotions are entwined with this.

The Machinehood Manifesto's definition of intelligence is almost identical to Hayles' definition of nonconscious cognition:

> "Intelligence is the ability to sense one's environment, follow a nonlinear set of rules, and adapt those rules based on the outcome of one's actions." (*Machinehood*, p 316)

Both Hayles' and the Machinehood's definitions are designed not to exclude AI, although Hayles separates cognition from thought and consciousness to deliberately avoid the concept of intelligence.

Emotion and care are *embodied* and closely connected to the *materiality* of the AIs in my sample of novels. The AIs use their physical bodies (whether a spaceship or a humanoid body) to sense the needs of those they care for and to assist them. They experience distress when their bodies do not allow them to care for their loved ones. The materiality of technology shapes what is possible (Wendy Chun 2008; 2017).

Care is *social* and depends on mutual emotional bonds between the AI and another person.

A theory of AI that is founded on care should therefore understand AIs as *embodied* (their physical means of sensing, interpreting and acting upon the world) and *social* (their relationships with others, and the social data they are trained on).

5. AI as embodied: sensing as care

Machine vision is often an aspect of how these AIs care for others. If AIs see, it must be through machine vision. But sight is not always emphasized. When it is, omnivoyance (see Liljefors, Noll, and Steuer 2019) is presented as an integral (or at least desired) part of their care, as for the benevolent AI that governs all of Earth in Shusterman's Thunderhead.

Sometimes their machine vision is faulty: overfitted so it's no longer trustworthy: "Every image processor we have is running cranked up far beyond sanity. We're overfitting on everything. Once a spider flagged the shadows of two trees as a threat, and I attacked it furiously for a few minutes. But better this than unaware." (*The Salvage Crew,* 204). (AI is fighting for itself but also its people.)

Others have humanoid or at least mobile bodies, but their minds tend to be distributed, weaving through data sources. Drones are common for enhancing vision – these AIs reject the singular point of view that limits unaugmented human vision.

6. Conclusion: why is this paper useful?

Fictional caring AIs provide a useful metaphor for thinking through how actual AIs are also social and embodied. Could thinking through the lens of care help us unravel the problem of bias in machine learning by seeing it as social and embodied?

How do we think about AI as social? Social datasets, certainly, but how do we understand mutuality and care in encounters between an AI and a human? Is our relationship with an AI, a corporation or capitalism itself?

Can or could AI *really* care? Do we want it to?

This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 771800).