



DESIGNING CARE AS AN AI SAFEGUARD: A VISUAL CONSIDERATION

Russell Suereth¹

¹*Salve Regina University*

Abstract

The problem this visual essay addresses is that safeguards are not designed in AI systems. One such safeguard is caring, a vital human characteristic that enables us to interact with others and the world around us in the right ways. The essay addresses the problem through a fictional comics story that targets general readers and AI designers. In the story, the main character is an AI system that wants to care. This AI system has found evidence of a caring foundation already in its design. However, more design work is needed for further caring to emerge. This visual essay hopes to show that comics can be a valuable tool for enabling AI designers to build safeguards such as caring in AI systems.

Keywords

Artificial Intelligence, AI, Safeguards, Caring, Comics

Prologue

This visual essay addresses the need for safeguards in AI designs. Designing safeguards in AI systems can be challenging because safeguards do not result in direct profits for AI corporations. This essay highlights the significance of safeguards, and particularly the safeguard of caring. However, much work is needed to design safeguards in AI, and this essay may be viewed as a beginning of that design work.

This research aims to show that a comics story can interest readers and be a call to action for designing caring in AI systems. The objectives of this research are:

- Identify aspects of storytelling in comics.
- Explore the purpose of the storyworld in comics.
- Investigate the concept of motivation and conflict in the comics story.
- Examine how comics can benefit our critical thinking.
- Consider how comics can offer new ways of understanding.
- Explore how comics can be a call for action.

This essay begins by describing the first scene of the comics story. The scene sets up the story by giving a short view of the storyworld. The essay describes each scene to help the reader understand how a comics story could be created to depict the issues that plague our world today. As Gannon and Fauchon remark, illustrative storytelling "requires much deliberation, planning and development, both intellectually and through creative practice" (Gannon & Fauchon, 2021, p. 39). AI, in general, is focused on artificial intelligence, which is a digital form of our human intelligence. This essay discusses AI through the medium of a comics story, a human mode of communication that is valuable for describing human things.

Scene One: The Storyworld

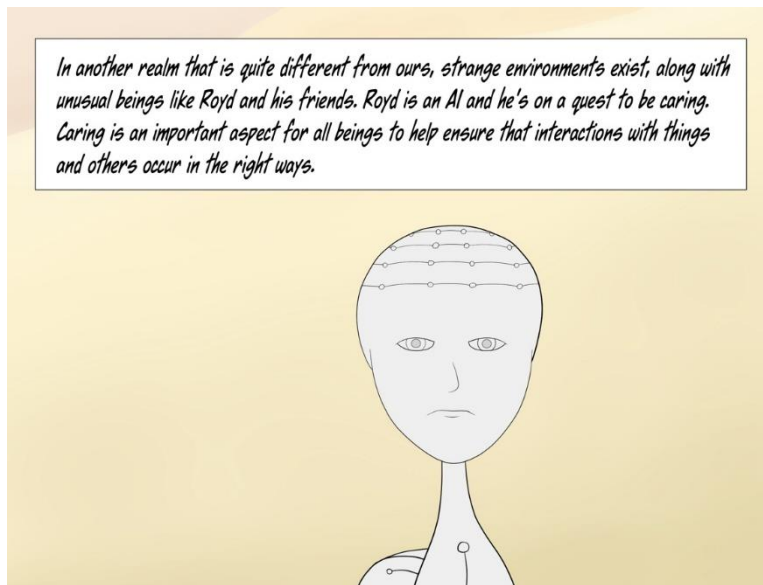


Figure 1. The storyworld. Digital drawing by the author.

This first scene briefly depicts the storyworld in this comics story. The world depicted in figure 1 is unusual because it contains strange environments and characters not from our realm. A depiction of the storyworld is essential in comics because it lays a foundation for the reality or non-reality the reader should expect. According to Gannon and Fauchon, “all narratives exist within a unique storyworld that describes the wider environment that the story is situated within....Even the most surreal or fantastical narrative situation will operate with an internal logic” (2021, pp. 44-45).

In this scene, the main character, Royd, is introduced in a desert-like setting. Here, the desert seems void of color, or in other words, without the caring that Royd is seeking. Royd looks serious or concerned in this panel, reflecting his

conflicts described in later story panels.

An Alternative Reality Opens Perspectives

Future worlds and alternative realities depicted through illustration unblock readers' perspectives. This depiction enables readers to see that change is needed and possible. According to Gannon and Fauchon (2021, p. 211), open perspectives create pathways for authors to enable change.

In this visual essay, the change is the design of safeguards in AI systems. Safeguards are needed to help ensure that AI systems will not go beyond their intended use and harm humans (Bostrom, 2014, pp. 289, 318). These safeguards arise by designing caring into AI systems (Suereth, 2024, pp. 1-2). Caring has been significant in our development throughout our human history (Hodges, 2017, pp. 33, 35, 44). According to the philosopher Martin Heidegger, caring is a crucial and intrinsic aspect of being human (Heidegger, 2010, p. 300). Today, caring is vital in our homes, work, and hospitals.

Scene Two: The Beginning



Figure 2. The beginning. Digital drawing by the author

This scene is the beginning of the dialogue between the story's characters. Figure 2 is another colorless desert scene that again seems void of emotion or caring. In the background, the mountains of humanity are visible yet seem slumbering.

In the scene, Royd is talking with his friend Pierre. The dialogue in these panels is short to fit the space in this visual essay. Dialogue in a complete comics would have more space to breathe and thus be more realistic. Stan Lee remarks:

Remember, very rarely in life do people simply come right out and say what they're really thinking. More often, they're oblique or guarded about their motives. Beware of ham-fisted diatribes where your hero or villain just comes right out and says exactly who they are and what they want. (Lee, 2011, p. 193)

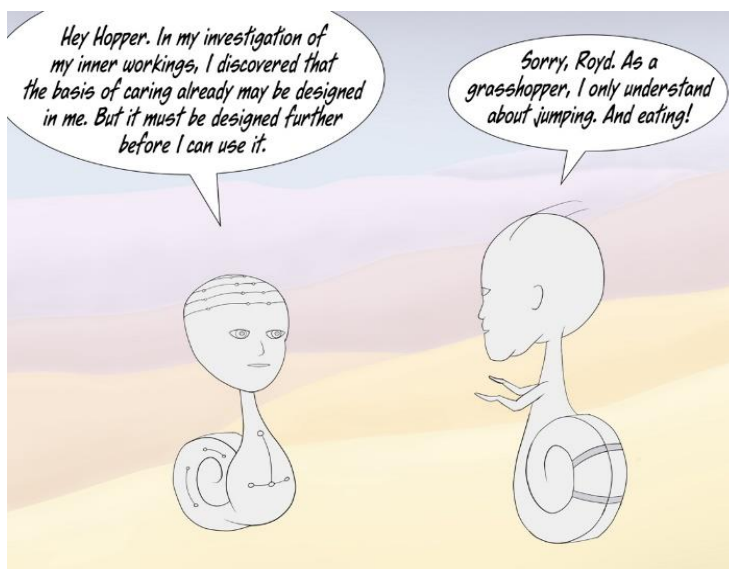
Motivation and Conflict

In this dialogue, we learn what Royd seeks — the capability to care. This quest is the story's primary motivation, and establishing this focus is crucial to any story. Kneece remarks:

Whatever the main character wants becomes what your story is about. What the character wants is the character's motivation. If a character doesn't want anything in particular, then the character's motivations are poor, and the story itself will suffer. (Kneece, 2015, p. 185)

Conflict is also important in a story, and Royd experiences conflict in this scene when Pierre claims that he knows nothing about caring. According to Scott McCloud, some conflicts appear to be external, "track them to their source, though, and nearly all conflicts are internal – because they all start with someone, somewhere, wanting something" (McCloud, 2006, pp. 66-67). Although this comics story is too short for an internal analysis of Royd, an investigation of his background and character would be intriguing in a larger work.

Scene Three: The Journey



In this scene, Royd tries to tell his friend Hopper about his quest. Royd understands his inner design, as shown in his explanation to Hopper. However, Hopper does not seem interested. Royd's understanding contrasts with Hopper's lack of understanding, and this difference may create tension between the two characters.

The setting of this scene in figure 3 is similar to the setting in scene one. However, this scene focuses on the background's colorful features. These colors reflect an emerging aspect of Royd. Royd's concern and quest for caring suggest he already has a small degree of caring.

Figure 3. The journey. Digital drawing by the author

Comics Enable Understanding and Critical Thinking

Comics engage us in a variety of ways. For Eisner, the combination of words and illustrations in comics represents a different way of reading and experiencing. "When this language is employed as a conveyance of ideas and information, it separates itself from mindless visual entertainment. This makes comics a storytelling medium" (Eisner, 2008, p. 33). Some comic stories focus more on the illustration than the textual dialogue and narrative. For Eisner, in these comics, "the art becomes the story, as in tapestries" (Eisner, 2008, pp. 83-84).

Comics can be a valuable medium for learning and understanding. In today's multi-modal world, students learn in multiple ways, and the variety of modes employed when reading comics is beneficial (Afrilyasanti & Basthomi, 2011, pp. 558-559). Studies have also shown that comics enable students to appreciate teachers and encourage reflection and critical thinking (Adnan et al., 2019, pp. 131-132). Furthermore, using comic books in the classroom can improve critical thinking in students (Damopolii, Paiki, & Nunaki, 2022, p. 354).

Scene Four: Turmoil

Here, Royd expresses his challenge to improve his caring capabilities and operations. The setting in figure 4 may seem quite unusual compared to the environments in our physical world. The blue colors suggest the funk that Royd is in, and the waves indicate his turmoil. According to Gannon and Fauchon, color schemes are valuable in illustrations and can communicate different concepts and feelings (2021, p. 45).

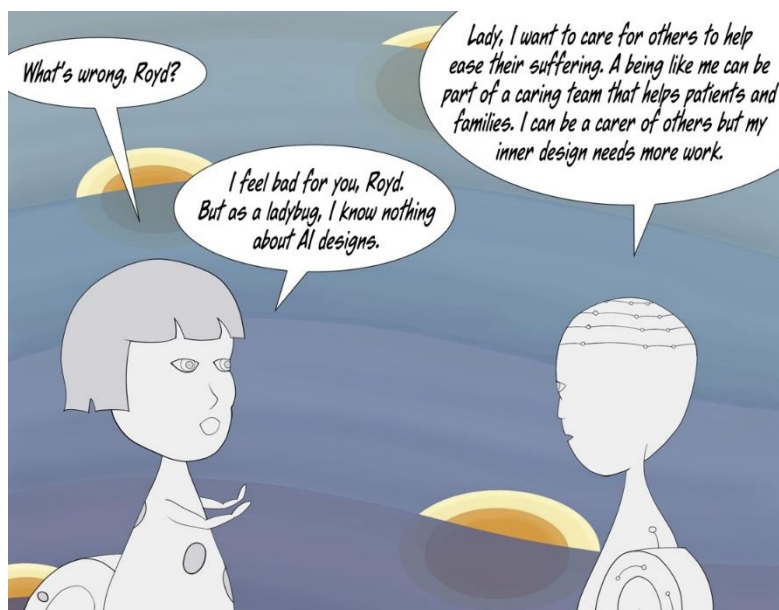


Figure 4. Turmoil. Digital drawing by the author

Comics and New Ways of Understanding

The concepts of caring in this visual essay are relatively simple. However, concepts of caring can be more complex and philosophically oriented. These more complex concepts should not deter us from using the comics medium. As Cerić and Cerić note, comics provide new ways of understanding because both visual and verbal skills are employed in reading them (Cerić & Cerić, 2022, p. 3). The authors suggest that comics can create meanings in unique ways due to their twined connections of words and illustrations (Cerić & Cerić, 2022, p. 3). Because they combine words and illustrations, comics can more vividly represent philosophical concepts (Cerić & Cerić, 2022, p. 10).

Scene Five: Resolution

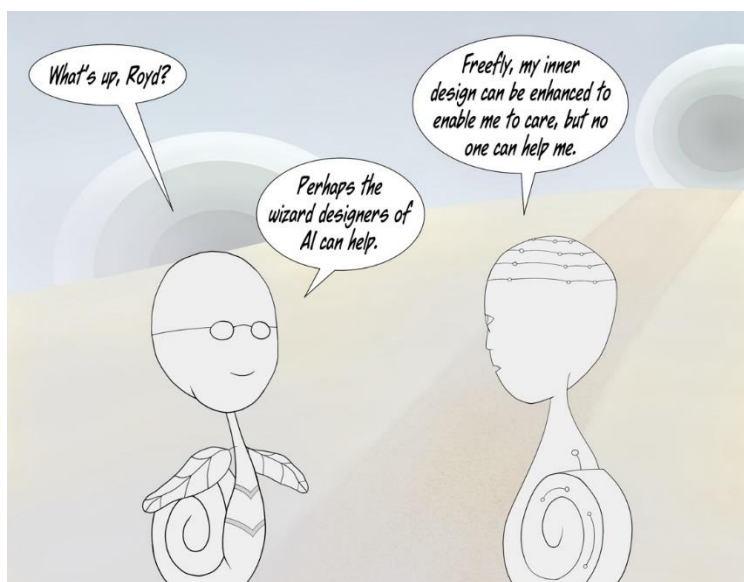


Figure 5. Resolution. Digital drawing by the author

In this scene, Royd discusses his tribulations with Freefly. The background setting in figure 5 is hazy and indicates Royd's direction challenge. The watching eyes highlight the increased observance of Royd's plight. The pathway suggests an avenue toward resolution. The last panel in this scene, figure 6, suggests a course forward.

Comics and Activism

In contrast to the prior scenes, this scene suggests action. Interestingly, comics can offer a viable medium for action. As Gannon and Fauchon note, graphic imagery works well in activist movements that require clear and urgent communications (2021, pp. 198-199). Nijdam remarks that mainstream comics advocate social justice (Nijdam, 2020, p. 204). Accordingly, due to

its ability to provide a clear and urgent message, comics can be a viable medium for issues related to social, human, and non-human concerns.

Epilogue

In this visual essay, the illustrations and dialogue were created first as a foundation for the tale of AI safeguards. Then, the narrative was written to describe the comics in the context of human or non-human change. However, as the visual essay progressed, the comics and narrative informed one another, which led to iterations of modifications for both. Indeed, these illustrations are not meant to be impeccable samples of illustrations or comics in general.

In this essay, the comics story is an example of a simple introduction. An actual story could expand this example by taking paths that lead to conflicts, successes, and new realizations. Hopefully, this story introduction may help readers see how their own comic story about AI or any social issue could begin.

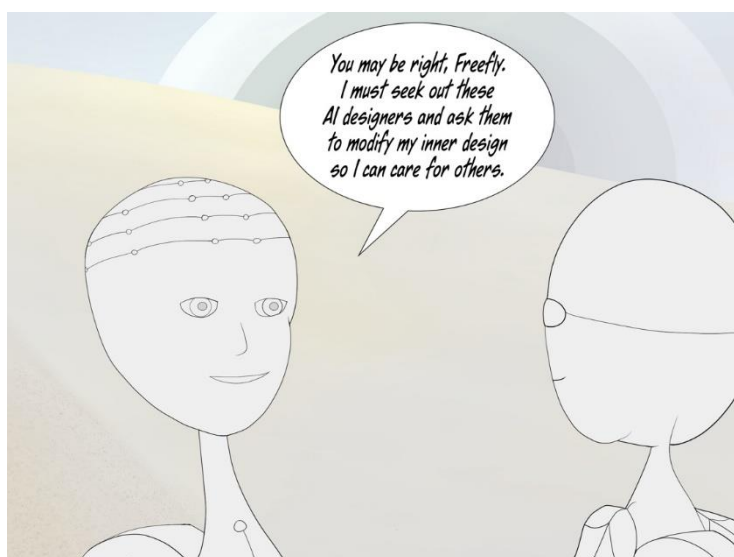


Figure 6. Activism. Digital drawing by the author

References

- Adnan, M., Abdullah, J.F., Ibharim, L. F. M., Hoe, T. W., Janan, D., Abdullah, N., & Idris, N. (2019). Expanding opportunities for science, technology, engineering and mathematics subjects teaching and learning: Connecting through comics. *Malaysian Journal of Medical Sciences*, 26(4), 127–133. doi:10.21315/mjms2019.26.4.15
- Afrilyasanti, R., & Basthomi, Y. (2011). Adapting comics and cartoons to develop 21st century learners. *Language in India*, 11(11), 552–568.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Cerić, H., & Cerić, E. (2022). A comicsophy approach to teaching philosophy. *Childhood & Philosophy*, 18, 1–22. doi:10.12957/childphilo.2022.64892
- Damopolii, I., Paiki, F. F., & Nunaki, J. H. (2022). The development of comic book as marker of augmented reality to raise students' critical thinking. *TEM Journal*, 11(1), 348–355. doi:10.18421/TEM111-44
- Eisner, W. (2008). *Graphic storytelling and visual narrative: Principles and practices from the legendary cartoonist*. W.W. Norton & Company.
- Gannon, R., & Fauchon, M. (2021). *Illustration research methods*. Bloomsbury Publishing.
- Heidegger, M. (2010). *Being and time* (J. Stambaugh, Trans.). State University of New York Press.
- Hodges, B. H. (2017). Carrying, caring, and conversing: Constraints on the emergence of cooperation, conformity, and language. *Interaction Studies: Social Behavior and Communication in Biological and Artificial Systems*, 18(1), 26–54. doi:10.1075/is.18.1.02hod
- Kneece, M. (2015). *The art of comic book writing: The definitive guide to outlining, scripting, and pitching Your sequential art stories*. Watson-Guptill Publications.
- Lee, S. (2011). *Stan Lee's how to write comics!* Watson-Guptill Publications.
- McCloud, S. (2006). *Making comics: Storytelling secrets of comics, manga and graphic novels*. William Morrow.
- Nijdam, E. (2020). Introduction: The social justice work of German comics and graphic literature. *Seminar, A Journal of Germanic Studies* 56(3–4), 191–211. doi:10.3138/seminar.56.3-4.intro.
- Pratt, H. J. (2009). Narrative in comics. *The Journal of Aesthetics and Art Criticism* 67(1), 107–117.
- Suereth, R. (2024). Considering caring as a safeguard in artificial intelligence. *New Techno Humanities*. <https://doi.org/10.1016/j.techum.2024.01.002>