# **Final Draft in HCC**

"**Why humans can’t trust AI: You don’t know how it works, what it’s going to do or whether it’ll serve your interests?"**

<https://theconversation.com/why-humans-cant-trust-ai-you-dont-know-how-it-works-what-its-going-to-do-or-whether-itll-serve-your-interests-213115>

**A graph with text and numbers

Description automatically generated**

**Figure 1. Humans trust on Artificial Intelligence and humans**

**Summary of the article:**

Artificial intelligence (AI) is currently influencing every aspect of our daily lives, driving everything from credit reports to facial recognition. However, because of their unpredictability and the opacity of their inner workings, trusting modern AI systems is a huge difficulty. Because predictability and trust are fundamentally interconnected, trust is damaged when artificial intelligence performs unexpectedly (Reinhardt, K. (2023)). As artificial intelligence (AI) systems learn from data, they modify parameters. These systems are commonly based on deep learning neural networks, which simulate the human brain. The AI explainability challenge is brought about by the fact that although they are good at making predictions, the reasoning behind their choices can often be unclear. Particularly problematic is this lack of transparency in high-stakes scenarios, such autonomous cars making moral decisions.

The AI alignment challenge refers to how difficult it is for AI to align judgments with human expectations because, unlike humans, AI does not change its behavior depending on ethical principles. Human participation in AI decision-making is recommended to increase trust in the technology, provided that humans are informed or involved in the process. But as AI grows more complex and is included into vital systems, it becomes increasingly important to address explainability and alignment concerns in order to preserve credibility. Because AI systems are unpredictable and lack human understanding, humans find it difficult to trust them (Reinhardt, K. (2023)). To ensure that AI systems are worthy of our trust in the future, further research is required to solve these trust difficulties and make them more predictable and in line with human norms.

## **Relation with HCC:** The article is strongly related to Human-Centered Computing (HCC) due to its emphasis on trust and usability in artificial intelligence (AI) systems. HCC places a heavy emphasis on ethical issues, user-centric design, transparency, and usability goals. To develop confidence in AI, the article emphasizes the need to match AI decisions with ethical standards, encourage user-centric design, and increase transparency. Additionally, the article touches on concepts like the Gulf of Execution and Gulf of Evaluation, emphasizing the importance of bridging the gaps between users' mental models and AI behavior. Overall, the article's discussion aligns with the principles and objectives of HCC, emphasizing the importance of trust and usability in AI systems. As a whole, the article's topic, which emphasizes the value of trust and usability in AI systems, is consistent with the goals and concepts of HCC.

**Interesting and Important points:**

The article explores the complex relationship that exists between artificial intelligence (AI) systems and trust, revealing a crucial topic. It highlights how difficult it is to trust AI because of its unpredictable nature, which is mostly caused by the "black box" problem—a lack of openness in AI decision-making(Kucharzyk, B. (2023)). The article emphasizes how unpredictability reduces trustworthiness, making predictability essential to establishing trust. The AI explainability challenge is raised by the behavior of deep learning neural networks, which include billions of parameters.

Humans find it challenging to trust AI when its decision-making lacks transparency and the ability to rationalize its choices, especially in high-stakes scenarios such as autonomous vehicles making moral decisions. This is a really important topic that needs to be further investigated. As AI is incorporated into more and more important systems, trust in these systems is essential. The article highlights the necessity to resolve trustworthiness concerns before they become insurmountable. For AI technology to advance, it is essential that we investigate ways to make it more consistent in accordance with human ideals.

**Examples for concern with trust in AI:**

Autonomous Vehicles and Safety Concerns:

Challenges: Trust issues with safety and critical scenario decision-making arise with autonomous cars.

User-Centered Design Approach: To address safety concerns, incorporating user feedback into the development process. Creating user interfaces that effectively convey the decisions and actions made by the vehicle. Test with passengers to see how they respond to various situations so that you can adjust the AI's behavior. Putting a strong emphasis on educating users about the limitations and possibilities of autonomous systems.

A graph showing how self-driving cars are being made

Description automatically generated

**Figure 2. humans trust on autonomous cars**

Voice Assistants and Privacy Concerns:

Challenge: Due to their constant listening and data storage, voice assistants like Siri or Alexa frequently cause privacy concerns.

User-Centered Design Approach: By connecting language to data operations and facilitating logical comprehension, provide an intuitive privacy interface. Overcome memory constraints with brief, memorable privacy option descriptions. Increase attention by using visual cues and notifications for policy changes. Optimize Fitts’s Law for easy interaction, with clickable features for efficient privacy preference adjustment.

## **Good or bad user centric design:**

The concepts of user-centric design play an important role in determining the reliability of artificial intelligence (AI) systems. Designers need to prioritize transparency in order to build trust, by designing user interfaces that offer valuable insights into AI's decision-making procedures (Zimmermann, T. (2023)). Even if the technical details are beyond the user's capacity to understand, visualization tools and straightforward explanations can help the user experience in control and confident. Furthermore, feedback mechanisms are critical for keeping people informed about AI-generated decisions in a clear and timely manner. It is important to acknowledge the ethical implications of AI systems and make sure they are ethically programmed to align with human ideals.

Another important component is explainability, since despite the underlying complexity, interfaces should aim to deliver AI conclusions in ways that are easy for users to understand. Furthermore, human control is critical in empowering and developing confidence by allowing users to challenge, oppose, or modify AI-generated judgments. On the other hand, designers should stay away from actions that decrease confidence. Building confidence in AI systems can be greatly aided by user-centric design, which places a high priority on transparency, feedback, ethics, explainability, and user control while avoiding some common errors. This will make AI systems more dependable, predictable, and trustworthy.

## **Personal Reflection or suggestions:**

The core issue revolves around AI's unpredictability due to its inner workings being largely impenetrable, leading to a lack of transparency and explainability. The following usability goals and design principles suggest improvements for people to trust AI.

1. Effectiveness: The degree to which AI fulfills its intended roles affects confidence in the technology(Kleinsmith, A. “Week 2 Lecture Slides”). If AI's decisions are unpredictable or unexplained, it limits effectiveness and erodes trust.

2. Efficiency: AI systems should assist users in performing activities efficiently(Kleinsmith, A. “Week 2 Lecture Slides”). AI behavior that is unpredictable might reduce work efficiency and weaken trust.

3. Security: AI should keep people safe from potentially dangerous situations(Kleinsmith, A. “Week 2 Lecture Slides”). The unpredictability of AI decisions might lead to safety issues and a lack of confidence.

4. Utility: People have faith in AI when it offers the appropriate features(Kleinsmith, A. “Week 2 Lecture Slides”). Unpredictable AI behavior can reduce the utility of AI and decrease users trust.

5. Learnability: Users will find it more difficult to trust and learn AI systems if their behavior is unexpected, as they will not be able to predict it(Kleinsmith, A. “Week 2 Lecture Slides”). Therefore, it is important to design them such that consumers may simply understand them.

6. Memorability: A predictable and consistent AI system is easier to remember and trust(Kleinsmith, A. “Week 2 Lecture Slides”). In contrast, unpredictable AI systems are challenging to memorize and, consequently, trust.

When considering how AI interfaces should make their capabilities and activities evident to users, the idea of "affordances and signifiers" is important(Kleinsmith, A. “Week 4 Lecture Slides”). Users should be able to predict how AI will respond, and a lack of clear affordances or signifiers leads to unpredictability and decreases trust. The topic on the "Gulf of Execution" and "Gulf of Evaluation" is relevant, as the unpredictability of AI decisions creates a considerable gap between users' mental models and the AI's behavior (Kleinsmith, A. “Week 5 Lecture Slides”). Establishing trust in AI systems requires bridging these gaps between gulfs.

In conclusion, the design and usability of AI systems have a direct impact on trust amongusers. The core principles of usability are challenged when AI performs in an unpredictable manner, making it challenging for users to rely on and trust these systems. Developing a connection between the internal workings of AI and human expectations is crucial to building trust in the system.

**Storyboard supporting the suggestions:**

**A diagram of a diagram

Description automatically generated**

**Figure 3. Storyboard showing the factors that can improve humans trust on AI.**

**References:**

1.Reinhardt, K. (2022). Trust and trustworthiness in AI ethics. *AI And Ethics*, *3*(3), 735–744. <https://doi.org/10.1007/s43681-022-00200-5>

2. Kaplan, A. D., Kessler, T. T., Brill, J. C., & Hancock, P. A. (2021). Trust in Artificial Intelligence: Meta-Analytic Findings. *Human Factors*, *65*(2), 337–359. <https://doi.org/10.1177/00187208211013988>

3. Wang, R. (2023, May 18). *Investigating and designing for trust in AI-powered code generation tools*. arXiv.org. <https://arxiv.org/abs/2305.11248>

4. Brożek, B., Furman, M., Jakubiec, M., & Kucharzyk, B. (2023). The black box problem revisited. Real and imaginary challenges for automated legal decision making. *Artificial Intelligence and Law*. <https://doi.org/10.1007/s10506-023-09356-9>

5. Independent high-level expert group on artificial intelligence - BSA. (n.d.). <https://www.ai.bsa.org/wp-content/uploads/2019/09/AIHLEG_EthicsGuidelinesforTrustworthyAI-ENpdf.pdf>

6. “Week 2 Lecture Slides”, Kleinsmith, A. “Week 3 Lecture Slides”, Kleinsmith, A. “Week 5 Lecture Slides”, Kleinsmith, A.

7. Richter, F. (2018, March 20). *Infographic: Consumer Concerns about Self-Driving Cars*. Statista Daily Data. [https://www.statista.com/chart/5950/concerns-about-self-driving-cars/](file:///Users/vineethpetnakota/Desktop/Richter,%20F.%20(2018,%20March%2020).%20Infographic:%20Consumer%20Concerns%20about%20Self-Driving%20Cars.%20Statista%20Daily%20Data.%20https:/www.statista.com/chart/5950/concerns-about-self-driving-cars)

8. Williams, R. (n.d.). *Research IntelligenceR: Consumers don’t trust AI to replace human interaction just yet*. <https://www.mediapost.com/publications/article/387209/consumers-dont-trust-ai-to-replace-human-interact.html>