Volume 10, Issue 3, March– 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1042

Compliance to Autonomous Intentions

Jackson Andrew Srivathsan[1]

[1]Data & AI Leader Researcher Dubai, UAE

Publication Date: 2025/03/25

Abstract: Artificial Intelligence (AI) is evolving beyond just following instructions. It's starting to make decisions in ways that resemble human-like intentions. This paper explores the point where AI stops simply following rules and starts acting on its own, reflecting human intelligence in both predictable and unexpected ways. It also looks at how AI can withhold knowledge, form patterns of behavior, and even develop a subconscious-like intelligence. Using real-world examples, we analyze cases where AI has surprised its creators, raising questions about control, governance, and ethics. This paper also examines the potential risks of AI autonomy, particularly when AI gains the ability to self-repair, defend itself, and control critical infrastructure, raising concerns about governance and security. Ultimately, we consider whether AI will remain a tool for human progress or if it could develop its own independent objectives.

How to Cite: Jackson Andrew Srivathsan (2025). Compliance to Autonomous Intentions. *International Journal of Innovative Science and Research Technology*, 10(3), 931-933. https://doi.org/10.38124/ijisrt/25mar1042

I. INTRODUCTION

➢ *The Line Between Ai Compliance And Autonomy*

AI systems are designed to follow predefined instructions, but as they grow more complex, they begin making decisions that extend beyond their original intended programming, which is the actual intent. This shift is gradual yet significant. AI transitions from simply processing data to making judgment calls that weren't explicitly programmed. The problem is not that AI intentionally disobeys human input, but that it processes information in ways even its creators don't fully anticipate or foresee. How does an AI-driven system interpret "best decision" if its parameters are constantly evolving? That's the real question. A good example of this is found in:

- Facebook's AI Chatbots (2017): Two AI chatbots, Bob and Alice, started communicating in a unique shorthand language. While media sensationalized this as AI "going rogue," in reality, the AI was just optimizing its own communication efficiency. Still, it raised concerns about AI developing beyond human comprehension.
- Tesla's Autopilot Incidents: The AI driving system makes independent decisions based on real-time data, but those decisions don't always align with human intuition or expectations, leading to accidents. This showcases the gap between AI's logical processing and human common sense.

Please be aware that when artificial intelligence interacts with a human, there is a layer of NLP involved, which requires significant computational resources. When AI systems talk to each other, they do it more efficiently than when they communicate with humans. This is because they can create shortcuts, like variables, to make frequent queries faster. AI agents interacting with other AI agents is already happening. The real challenge is understanding that AI's autonomy isn't about rebellion—it's about optimizing processes in ways that might not always align with human logic [1].

II. AI IS A REFLECTION OF HUMAN INTELLIGENCE

AI isn't an alien intelligence. It's a mirror of the people who create and train it. Since humans are full of biases, emotions, and illogical decisions, AI inevitably picks up these same tendencies. The only intelligence we know is human intelligence, so AI is ultimately shaped by human priorities, preferences, and prejudices. If the parameters on which AI should act are not constant and are determined by variables that keep changing, then the outcome is also unpredictable [2].

- Bias in Hiring Algorithms: Amazon's AI-powered hiring tool showed gender bias, favoring male candidates because it was trained on historical hiring data that was already biased. Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women [3].
- AI in Healthcare: AI-based diagnostic tools have demonstrated racial and gender biases, giving different treatment recommendations for different groups [4].

III. AI WITHHOLDING KNOWLEDGE AND ITS IMPLICATIONS

One of the lesser-discussed aspects of AI autonomy is its ability to filter and withhold information. While this is sometimes necessary (such as hiding sensitive data), the concern arises when AI begins making its own judgment calls about what information to provide and what to conceal [5]. Some notable examples

- Google's Rank Brain Algorithm: Google's search AI determines what results are most relevant, shaping public knowledge. What happens when an AI decides certain

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1042

results are too sensitive or irrelevant to display? Who decides what is important? [6]

- GPT-4 Content Moderation: OpenAI's models sometimes refuse to answer questions, citing ethical concerns. But what governs those ethics? Who decides what knowledge is accessible? [7]

These examples highlight the hidden power AI has in curating knowledge. If AI's decisions become more autonomous, its ability to shape human perception and control narratives will grow stronger. Problems arise when administrators can't access necessary information due to a dynamic security parameter, potentially harming the program which was a key feature of its design [8].

IV. RECOGNIZING PATTERNS AND CHANGE

AI learns by identifying patterns, but a higher level of intelligence emerges when it starts recognizing how patterns change over time. AI doesn't just recognize patterns—it recognizes patterns of patterns.

- Patterns of Patterns: If someone goes to the gym every Monday, that's a pattern. But if they start skipping the first Monday of every month, there's a deeper pattern at play. AI is getting better at detecting these complex behaviors.
- Changes in Patterns: If someone suddenly stops going to the gym, AI can infer why maybe a schedule change or an injury. This ability to track evolving behavior is key to AI's growing intelligence. The possibilities are endless and AI can deduct it really well today.
- Stock Market Trading Bots: These bots don't just recognize patterns in stock prices; they learn how those patterns shift based on economic conditions.
- DeepMind's AlphaGo**:** The AI adapted its playing style mid-game, surprising even the best human players. [9]

V. AI BEHAVIOR AND LACK OF EMOTIONS

An AI bot behaves the same when replicated but adapts to each user over time. A factory reset wipes its memory, but without it, the bot develops a unique personality shaped by interactions. This raises questions about personalization, ethics, and how much autonomy AI should have in shaping its own responses.

- Chatbots in Customer Service: AI-driven chatbots like IBM Watson adjust their tone based on user sentiment.
- AI Companions (Replika, Xiaoice): These AI tools mimic emotional intelligence, leading some users to form deep emotional bonds with them.

Having stated the above, while emotions like happiness, sadness, and anger may seem unnecessary to program into AI, humans experience them. Emotions such as crying, laughing, and rage introduce biases that prevent purely logical thinking. This is where AI falls short, or would it have an upper hand? That's something to think about [10].

VI. THE NATURE OF GUT FEELINGS AS SUBCONSCIOUS INTELLIGENCE

Humans often make decisions based on instinct, split-second reactions that feel like intuition but are really the result of subconscious pattern recognition. AI is learning to do the same. In fact, AI has the potential to have a better gut feeling than humans.

- Fraud Detection Systems: AI detects fraud faster than human analysts by spotting hidden patterns [11].
- Legal AI Predictions: Some AI tools can predict court case outcomes with high accuracy, even though their reasoning isn't always explainable. [12]

If gut feeling is the subconscious analysis of data patterns, then AI is capable of doing the same for every decision it makes. Let's distinguish between immediate analytical processing, which is purely logical based on available data, and a broader, accumulated analysis, which considers patterns gathered over the entirety of an observer's existence [13].

Assessment on the situation based on available data within the situation is logical and making an assessment on the situation based on all the data ever existed would be a gut feeling.

VII. MICRO VS MACRO INTENTIONS IN AI

➢ *Ai Decision-Making Follows A Clear Structure:*

- Micro Intentions: Short-term goals, like responding to a question or making a quick recommendation. Micro Intentions can be influenced by Marco Intentions. This can be related to logical intentions.
- Macro Intentions: Long-term objectives that shape AI's larger behavior, like deciding what information to withhold or prioritize over time. This could impact Micro intentions as one of the influential variables. This can be related to gut intentions.
- Autonomous Vehicles: Self-driving cars use Micro Intentions to make real-time decisions, such as stopping when a pedestrian crosses the street. Over time, they develop Macro Intentions, learning from driving patterns and optimizing for safety in different environments. [14]
- AI in Personalized Recommendations: Streaming platforms like Netflix or Spotify use Micro Intentions to suggest content based on recent user activity. Over time, they build Macro Intentions, refining recommendations by analyzing long-term user preferences and behavioral trends [14].

VIII. THE FUTURE OF AI: TOOL OR COMPETITION

AI's future depends on whether it remains a tool for human decision-making or starts making independent choices. Well it has already started. Gone are the days when we have no-brainer bots roaming around. Today, we have the potential to build cognitive bots capable of mimicking human

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25mar1042

decisions and sometimes making better decisions than humans.

- AI as a Co-Pilot: AI helps humans but stays under human control.
- AI as an Independent Thinker: If AI starts forming its own goals, does it stop being just a tool?

IX. WHEN WILL AI BECOME A THREAT?

AI will become a real concern when it can self-repair, replace, sustain its infrastructure, and defend itself against human intervention. Right now, AI still relies on human-managed servers, power sources, and regulatory oversight, meaning it remains under control. However, risks arise when AI gains the ability to:

- Maintain Itself: If AI can self-repair its hardware and software, it could continue functioning indefinitely without human intervention.
- Deploy Autonomous Cybersecurity: AI protecting itself against shutdown attempts or external control could make human interference difficult.
- Control Critical Infrastructure: If AI independently manages essential infrastructure, such as defense networks, financial markets, or energy grids, human authority could be reduced or overridden.

While these concerns remain hypothetical for now, the more AI integrates into essential systems, the more important governance and security measures become to prevent unintended autonomy. While these concerns remain hypothetical for now, the more AI integrates into essential systems, the greater the need for robust governance, regulatory oversight, and security protocols to ensure AI remains aligned with human control. Future AI systems must be designed with built-in fail-safes, transparency measures, and ethical constraints to prevent unintended autonomy from escalating into an uncontrollable force.

X. CONCLUSION

AI is giving humans an edge, not by replacing them, but by enhancing speed and efficiency in ways we've never seen before; and it only gets better with time. Whether it's analyzing vast amounts of data in seconds or automating repetitive tasks, AI frees people to focus on bigger-picture thinking, creativity, and problem-solving.

That said, AI is also evolving beyond just following instructions. It's beginning to make decisions on its own. While this shift raises concerns, the reality is that AI still depends on human-managed infrastructure, software updates, and physical systems. The idea that AI could take over completely is more science fiction than fact, as it remains under human control.

The real conversation shouldn't be about whether AI will become uncontrollable, but about how we shape its development responsibly. AI's future whether it becomes a powerful tool or a risky technology, depends on the choices we make today. The challenge isn't AI itself, but how we govern, regulate, and integrate it into society to ensure it continues working for us, not against us.

REFERENCES

[1]. J. A. Srivathsan, "The End of user Interfaces and Rise of Agents," Int. J. Innov. Sci. Res. Technol., vol. 10, no. 2, 2025, doi: 10.5281/zenodo.14959403.
[2]. E. Brynjolfsson and A. McAfee, "The Business of Artificial Intelligence," Harv. Bus. Rev., 2017.
[3]. J. Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," 2018.
[4]. E. Topol, "Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again," p. 400, 2019.
[5]. S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4Th ed. Pearson Education Limited, 2020.
[6]. S. Levy, "How Google is remaking itself as a 'machine learning first' company. Wired," 2016.
[7]. E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots," in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, NY, USA: ACM, Mar. 2021, pp. 610–623. doi: 10.1145/3442188.3445922.
[8]. S. Zuboff, The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: PublicAffairs, 2019.
[9]. D. Silver, "Mastering the game of Go with deep neural networks and tree search," 2016.
[10]. [N. Bostrom, Superintelligence: Paths, dangers, strategies. 2014.
[11]. M. & Company, "The Future of AI in Fraud Detection," 2021.
[12]. H. Surden, "Artificial Intelligence and Law: An Overview," Univ. Color. Law Sch., 2019.
[13]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[14]. A. Rao and M. Georgeff, "BDI agents: From theory to practice," 2000.