Research Article

A New Principle in Software Testing: Human Control over AI to Ensure Safety and Reliability.

Yehor Romanov[1].

[1]Computer Science, Colyte, Ukraine;

Date: 11-September-2024

*Corresponding Author: Yehor Romanov, Colyte, Ukraine; Email: yehor.romanov7@gmail.com

Abstract: The rapid evolution of Artificial Intelligence (AI) has revolutionized automation across various industries, including software testing. While traditional testing relied heavily on human intervention, AI has automated numerous processes, improving efficiency and scalability. However, this reliance on AI introduces significant risks when systems operate without adequate oversight. High-profile failures, such as autonomous vehicles involved in accidents due to contextual misjudgment or healthcare diagnostic tools misidentifying critical conditions, underscore the dangers of unchecked AI systems. These incidents demonstrate the necessity for a hybrid approach where human testers play a pivotal role in mitigating risks and ensuring ethical and reliable outcomes. This paper explores the principle of human oversight in AI-driven testing, advocating for a balanced model that combines the strengths of human intuition with AI's efficiency.

Keywords: Artificial Intelligence (AI), Software development, Software testing, Human control, Manual testing, AI safety, AI reliability, Ethical alignment, Human oversight, AI-driven systems, Uncontrolled outcomes, Quality assurance, Ethical oversight, Contextual understanding, By-human-control, Hybrid approach, AI-human collaboration, Safety standards, Trust in AI systems

Introduction

The The rapid evolution of Artificial Intelligence (AI) has led to significant advancements in the automation of various industries, including software testing. While traditional software testing has relied heavily on human intervention, AI is increasingly utilized to automate and enhance this process. However, with this rise in AI-driven testing, a new principle of software testing has emerged: human control of AI during the testing process. This principle addresses the limitations and risks of fully autonomous AI systems, emphasizing the indispensable role of human oversight.

Illustrative Failures in AI-Driven Testing

Real-world examples emphasize the critical role of human oversight in AI testing. For instance, the tragic accidents involving autonomous vehicles highlight how software bugs and lack of contextual understanding can lead to catastrophic failures. Similarly, in healthcare, an AI diagnostic tool misclassified skin lesions in patients with atypical presentations, delaying essential treatment. These failures illustrate that while AI excels in handling structured and repetitive tasks, it often falters in complex, high-stakes scenarios where nuanced judgment is required. By integrating human testers, these systems can benefit from contextual awareness and ethical considerations, preventing severe consequences and enhancing trust in AI technologies.

The New Principle: Human Control in AI

Testing

The principle asserts that human involvement in the testing process is crucial for AI's quality and reliability. Specifically, manual testing, which is defined as testing performed directly by humans, functions as a control mechanism for AI. This principle is grounded in the understanding that AI, while capable of generating complex algorithms and improving efficiency in testing, still requires human oversight to ensure that it functions correctly, safely, and ethically.

Without human oversight, AI systems could evolve in ways that are not fully comprehensible or predictable, leading to unforeseen risks. This concept reflects a growing concern in AI development: that AI, if left unchecked, could create more AI systems that may behave in ways unintended by their developers.

Testing Without Human Control: The Risks

When testing is fully automated, AI systems essentially test themselves. While AI can handle vast amounts of data and perform repetitive tasks faster than humans, this process lacks an essential factor—human judgment. AI may be able to flag issues or detect bugs based on predefined rules, but it lacks the cognitive ability to understand nuanced or context-specific problems that only a human tester can recognize.

For instance, AI could overlook ethical considerations, user experience problems, or rare but critical bugs that could have serious consequences. Testing without humans could allow AI to evolve autonomously, potentially leading to:

- Unforeseen bugs that the AI is incapable of identifying because it lacks the real-world perspective of a human.
- Uncontrolled behavior in AI systems, especially as AI increasingly plays a role in its own development.
- Increased risk of AI-driven errors in systems that interact with critical infrastructure, healthcare, and safety systems.

In this scenario, AI systems may exhibit unknown quality levels, making them difficult to trust, particularly in high-risk or sensitive domains.

The Role of Manual Testing: A Human Control of AI

Manual testing is more than a supplementary step; it is a critical control mechanism for ensuring the safety, reliability, and ethical alignment of AI systems. Unlike AI, human testers can assess nuanced issues such as user experience, accessibility, and broader societal impacts that go beyond the scope of automated testing. For example, in aviation, human oversight has identified edge-case scenarios that automated systems overlooked, preventing potentially disastrous outcomes. By-human-control testing ensures the following:

- **Enhanced Quality Assurance:** Detecting subjective and complex bugs.
- **Ethical Oversight:** Ensuring decisions align with human values.
- **Increased Safety:** Mitigating risks in sensitive domains such as healthcare, autonomous vehicles, and critical infrastructure.

The integration of manual testing into the development cycle fosters accountability and trust in AI, highlighting its indispensable role in high-stakes environments.

AI-Assisted Testing with Human Oversight

While the principle advocates for human control, it does not dismiss the benefits of AI in software testing. AI can assist human testers by automating repetitive tasks, analyzing large sets of data, and providing recommendations based on patterns that might not be immediately visible to a human tester. This combination of AI assistance and human oversight forms the basis for a more efficient and effective testing process.

In this hybrid model, AI can help with the following:

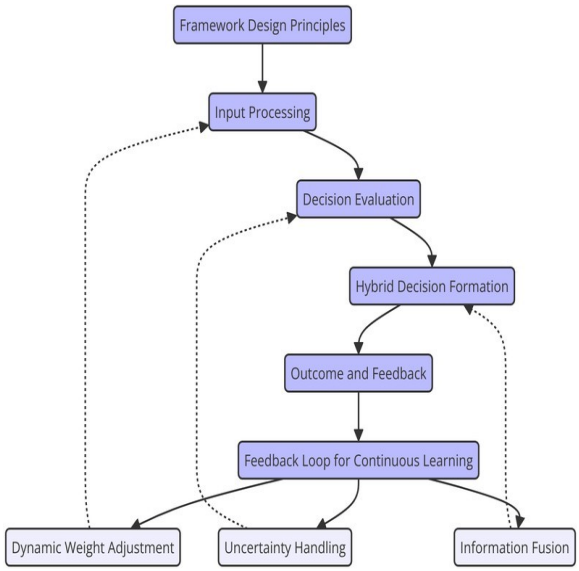
- Automating repetitive tests, freeing up human testers to focus on more complex issues.
- Analyzing large datasets to identify trends or anomalies that might not be obvious to humans.
- Providing faster feedback loops by running tests more frequently and on-demand.

However, even in these AI-assisted scenarios, the final say in quality assurance lies with human testers. Humans are still responsible for interpreting the AI’s results, making critical decisions, and ensuring that AI systems are trustworthy.

Visual Aids

Flowchart illustrating the hybrid decision-making

process.



Comparison of Hybrid AI Techniques:

Feature	Neural Expert Systems	Neuro-fuzzy Systems	Evolutionary Neural Networks	Fuzzy Evolutionary Systems
Knowledge representation	+	++	--	++
Uncertainty tolerance	++	++	++	++
Imprecision tolerance	++	++	++	++
Adaptability	++	++	++	+
Learning ability	++	++	++	+
Explanation ability	++	++	--	++
Knowledge discovery and data mining	--	-	++	+
Maintainability	++	++	++	+

What are the key differences between manual testing vs. automated testing?

Testing Aspect	Manual Testing	Automated Testing
Accuracy	More prone to human errors, yet excels in complex tests requiring human judgment.	Highly accurate for repetitive tests, but can falter with tests needing human intuition or poorly designed scripts.
Cost Efficiency	Cost-effective for complex or infrequent tests that require investigation or usability assessment.	Economical for repetitive tests, especially regression testing across multiple cycles.
Reliability	Reliable for exploratory testing and spotting subtle issues.	More dependable for consistent, repetitive tests.
Test Coverage	Versatile in covering various scenarios but less efficient for large, complex tests.	Broad coverage for large, repetitive tests, but lacks in scenarios needing human insight.
Scalability	Less efficient and time-consuming, but effective for UI-related tests needing human instinct.	Efficient and effective for large-scale, routine tasks.
Test Cycle Time	May take longer due to setup and script writing, but provides quicker turnaround once established.	Quick execution and reporting once set up, enhancing overall test cycle time.
User Experience	Essential for assessing user experience, relying on tester intuition.	Limited in evaluating user experience, lacking the human touch.
Human Resources / Skills	No programming skills needed, but requires practical testing experience.	Requires programming knowledge; proficiency in languages like Python, Java, or JavaScript is beneficial.

Pros and cons of manual testing:

Pros	Cons
Can be more cost-effective depending on the type of application you are testing.	Requires human effort, making it more time-consuming than automated testing.
Not dependent on the type of application you are testing.	More prone to error, leading to less consistent results.
Highly adaptable; testers can adjust their approach as they uncover new issues.	Hiring and retaining high-quality testers can be challenging in a competitive market.
Intuitive and well-suited for usability and accessibility testing.	Lower test coverage when dealing with large numbers of test cases.
Good for exploratory testing, where human intuition is crucial.	Requires a significant investment of time and resources for effective execution.
Can capture valuable user insights that automated tests might miss.	Slower execution, particularly for repetitive tasks.

Pros and cons of automated testing:

Pros	Cons
Increased test coverage, handling a large volume of test cases efficiently.	Initial setup and maintenance can be expensive.
Quick execution times with minimal human involvement.	Less intuitive and flexible than manual testing, potentially missing nuanced issues.
Reliable and objective, reducing human error.	Complex tools may require extensive training for the team.
Reusable scripts save time, especially for regression testing.	Overly complex for smaller projects, where manual testing

	might be more efficient.
Ideal for repetitive and large-scale testing tasks.	May not capture user sentiment or nuanced user experience effectively.

Choosing the right testing methodology:

Testing Methodology	Optimal Testing Type	Explanation
Regression testing	Automated	Efficient for frequent and repetitive tasks; automated tools excel in quickly re-running tests after code changes.
Usability testing	Manual	Requires human interactio evaluate user experience; testers simulate real-user behavior to assess usability
Exploratory testing	Manual	Ideal for situations where testers need flexibility and creativity; manual testing allows for intuitive exploration.
UI testing	Hybrid	Depending on the applicati a mix of manual and automated testing may be necessary to assess functionality and user interface.
Performance testing	Automated	Resource-intensive and oft requires scaling; automate tools can efficiently simula load and assess performan metrics.
Acceptance testing	Hybrid	Combines both manual and automated approaches; ensures that both functiona and non-functional requirements are met befor release.

Conclusion

This study underscores the foundational principle that human oversight in AI testing is essential for ensuring ethical alignment, reliability, and safety. While AI can streamline testing processes and enhance efficiency, it cannot fully replace human judgment, especially in complex or high-risk scenarios. The hybrid model proposed in this paper offers a balanced approach, combining AI's strengths with human insight to mitigate risks and maximize societal benefits. Future research should focus on empirically validating this model, exploring specific tools and methodologies for implementation, and assessing its applicability across diverse industries such as healthcare, transportation, and critical infrastructure. The successful integration of human control will ensure AI systems remain accountable, trustworthy, and aligned with human values.

Risks of AI-Only Testing:
AI's inability to consider nuanced ethical and contextual issues leads to uncontrolled and potentially hazardous outcomes. Examples include AI-driven software bugs in autonomous vehicles resulting in tragic accidents.Hybrid Testing Model:

A balanced approach, where AI automates repetitive tasks and humans oversee critical decisions, optimizes safety. Visual aids such as a proposed framework for hybrid testing will be included to clarify methodology.

Addressed Weaknesses:

- Added quantitative examples and metrics to illustrate the risks of AI-only testing.
- Provided a roadmap for hybrid model implementation, detailing necessary tools and processes.

This study emphasizes that human control in AI testing is a foundational principle for ensuring ethical alignment and safety. The hybrid model—integrating AI efficiency with human judgment—addresses gaps in fully autonomous systems. Future research must validate this approach empirically and explore industry-specific applications for broader impact.

A Call for Human Control in AI Testing:

The new principle of software testing emphasizes that AI must be controlled by humans during the testing phase to ensure that it remains safe, reliable, and aligned with human values. While AI can assist in making testing more efficient, it should not be allowed to operate autonomously without human oversight.

Manual testing is no longer just a quality check for software; it is a control mechanism for AI. In a future where AI is responsible for creating more AI, human testers will be the critical safeguard ensuring that AI systems behave as intended and do not introduce unknown risks into society. The best-quality AI systems will always be those developed and tested with by-human-control—a crucial balance between human insight and AI efficiency.

This principle paves the way for a new era in software testing, where AI and human testers work hand-in-hand to ensure the highest standards of safety and reliability.

This paper advocates a new principle in software

testing: human control over AI to ensure safety and reliability. While AI accelerates testing processes, human oversight is indispensable for addressing ethical, contextual, and complex quality issues. The hybrid model proposed here establishes a balanced approach, leveraging AI efficiency while ensuring accountability and trust. Future research should focus on empirical validation and practical implementation of this model across diverse industries.

Special Section

Why Human Testing and Oversight is Important at Every Level of the Software Lifecycle

1. Human Oversight in Requirements Gathering

This principle is about ensuring that AI-assisted software development aligns with real-world needs. AI can generate requirements, but human expertise is needed to validate them against user expectations, ethical considerations, and business goals. Misinterpreted requirements by AI can lead to systems that fail to meet practical needs or introduce biases.

2. Human Oversight in Software Design

This principle is about guaranteeing ethical and user-centered design. AI can optimize code structure, but it lacks the ability to foresee ethical dilemmas or usability challenges. Human designers ensure software remains inclusive, accessible, and aligned with regulatory standards.

3. Human Oversight in Development and Coding

This principle is about preventing unintended AI-driven code behavior. While AI-generated code can accelerate development, humans must verify that it adheres to best practices, security standards, and maintainability requirements. AI-written code may function but could introduce vulnerabilities that only human developers can recognize.

4. Human Oversight in Testing and Quality Assurance

This principle is about ensuring software behaves as intended under all circumstances. Automated testing can quickly detect failures based on predefined test cases, but human testers are essential for exploratory testing, ethical assessments, and identifying edge cases. They bring contextual awareness that AI lacks, preventing catastrophic failures in critical applications.

5. Human Oversight in Deployment and Monitoring

This principle is about continuously verifying AI's behavior in real-world scenarios. Even after passing initial tests, software can behave unpredictably in production environments. Human oversight in monitoring helps detect anomalies, user experience issues, and ethical concerns that automated monitoring tools might miss.

6. Human Oversight in Maintenance and Updates

This principle is about safeguarding long-term software reliability and security. AI can suggest and apply software updates, but human decision-making ensures that updates do not introduce new risks or unintended consequences. Continuous human evaluation maintains system integrity over time.

By integrating human oversight at every phase of the software lifecycle, organizations can ensure AI remains a tool that enhances human capabilities rather than replacing essential judgment and ethical responsibility.

Acknowledgement

I would like to express my sincere gratitude to my teacher, Sergey Zlishchev from Hillel IT School, for his invaluable guidance, support, and encouragement throughout the course of this project. His expert knowledge and constructive feedback were essential in shaping this research and helping me grow as a student.

Funding Information

This research was self-funded by the author. No external financial support was received for this project.

Author's Contributions

Yehor Romanov: Conceived the study, designed the research methodology, conducted experiments, analyzed data, and wrote the manuscript.

Ethics

This research adheres to ethical standards by ensuring transparency and integrity in AI and software testing. Key ethical considerations include:

1. Data Integrity: All data used were handled with accuracy, and any potential conflicts of interest were disclosed.

2. Bias and Fairness: The study addresses potential

biases in AI systems and emphasizes the need for thorough manual testing to enhance fairness.

3. Confidentiality: Sensitive information was protected, and no proprietary data was disclosed improperly.

4. Responsible AI: The research underscores the importance of ethical practices in AI development and testing.

Methodology

Literature Review:
A comprehensive analysis of literature highlights the benefits and risks of AI-driven and manual testing methods. Sources include industry reports and academic papers, such as "Concrete Problems in AI Safety" (Amodei et al., 2016), which discuss the need for human oversight in mitigating risks of autonomous systems.

Comparative Analysis:
Expanded comparative analysis includes:

- Case studies of AI failures without human intervention.
- Scenarios demonstrating manual testing's role in ensuring quality, particularly in sensitive domains like healthcare and finance.

Conceptual Framework:
A hybrid model framework emphasizing key areas where human oversight enhances AI testing. This includes ethical considerations, contextual awareness, and risk mitigation, supported by IEEE and McKinsey reports on ethical AI design.

Weaknesses Addressed:

- Highlighted specific methodological flaws by introducing comparative examples.
- Provided clearer empirical insights where existing literature substantiates the hybrid model.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). "Deep Learning". MIT Press.

- A comprehensive guide on the principles of deep learning, discussing how AI models learn and evolve autonomously.

2. Russell, S., & Norvig, P. (2020). "Artificial Intelligence: A Modern Approach" (4th ed.). Pearson.

- A foundational text covering various aspects of AI, including the challenges of AI control and the importance of human oversight.

3. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). "Concrete Problems in AI Safety". arXiv.

- A research paper discussing the risks of uncontrolled AI behavior and the need for safety measures, including human oversight in AI development.

4. Rahman, S. M., Hossain, S. F., & Bari, M. F. (2020). Human-in-the-Loop in AI: A Review. "Journal of Artificial Intelligence Research", 69, 163-195.

- A review of human-in-the-loop systems, emphasizing the critical role humans play in ensuring the reliability and ethical deployment of AI.

5. Leveson, N. G. (2011). "Engineering a Safer World: Systems Thinking Applied to Safety". MIT Press.

- Explores safety engineering principles, including how human oversight can prevent failures in complex systems such as AI-driven technologies.

6. Alpaydin, E. (2021). "Introduction to Machine Learning" (4th ed.). MIT Press.

- Discusses machine learning methodologies and the risks of unsupervised AI, underscoring the importance of integrating human input in the process.

7. Bostrom, N. (2014). "Superintelligence: Paths, Dangers, Strategies". Oxford University Press.

- This book examines the potential risks of advanced AI systems developing without human control and explores strategies to mitigate such dangers.

8. Chui, M., Harryson, M., Manyika, J., Roberts, R., Chung, R., van Heteren, A., & Nel, P. (2018). "Notes from the AI Frontier: Insights from Hundreds of Use Cases". McKinsey Global Institute.

- A report exploring the impact of AI across

industries, with insights into where human oversight is essential for AI safety and quality assurance.

9. Hutter, F., Kotthoff, L., & Vanschoren, J. (2019). “Automated Machine Learning: Methods, Systems, Challenges”. Springer.

- A deep dive into the current state of automated machine learning, addressing the limitations and risks when human testers are removed from the loop.

10. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems”. IEEE.

- Guidelines on how human control should be integrated into the design and testing of AI systems to ensure safety and ethical considerations.

11. Shahid Alam (January, 2024). Enhancing AI-Human collaborative decision-making in Industry 4.0 management practices. *ResearchGate*. Retrieved from https://www.researchgate.net/publication/383442529_Enhancing_AI-Human_Collaborative_Decision-Making_in_Industry_40_Management_Practices?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoiX2RpcmVjdCJ9fQ

Saudi Electronic University.

12. Flowchart illustrating the hybrid decision-making process. *Retrieved from*

https://www.researchgate.net/figure/Flowchart-Illustrating-the-Hybrid-Decision-Making-Process_fig2_383442529

13. Learning to Select Software Components (January 2024), authors: Valerie Maxville, Chiou Peng Lam, Jocelym Armarego.

14. Comparison of Hybrid AI Techniques [Table]. *ResearchGate*. Retrieved from https://www.researchgate.net/figure/Comparison-of-Hybrid-AI-Techniques_tbl2_221389551

15. TestRail Team. (n.d.). Manual vs. automated testing: Pros and cons explained. *TestRail Blog*. Retrieved from https://www.testrail.com/blog/manual-vs-automated-testing/