

Why Machine Ethics ?



Presented By
Archana Kumari
0190443310

Outline

- Introduction to Intelligent Machines.
- What is Ethics ?
- Machine Ethics
- Why machine ethics?
- Role of Ethics
- Conclusion
- Summary

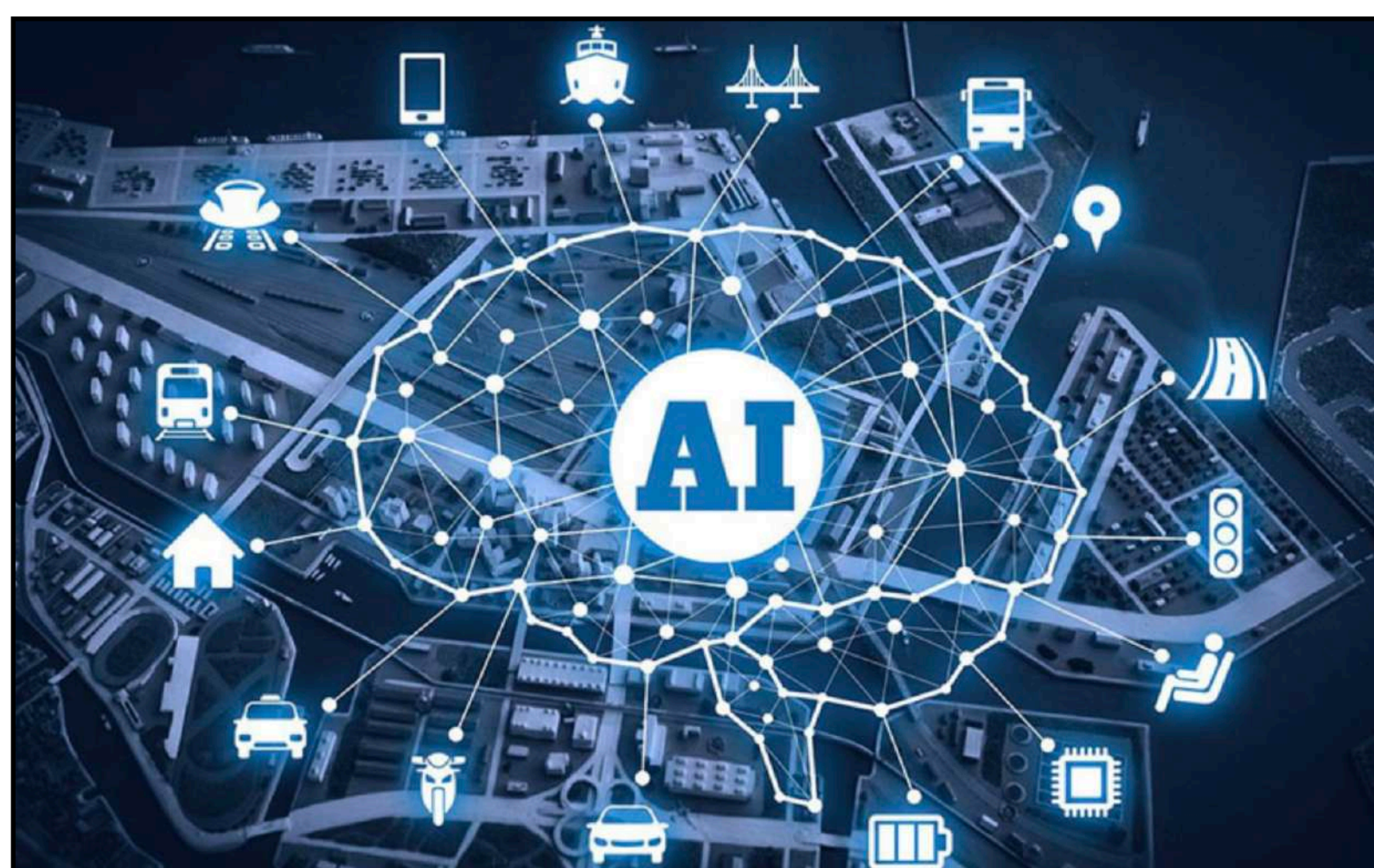
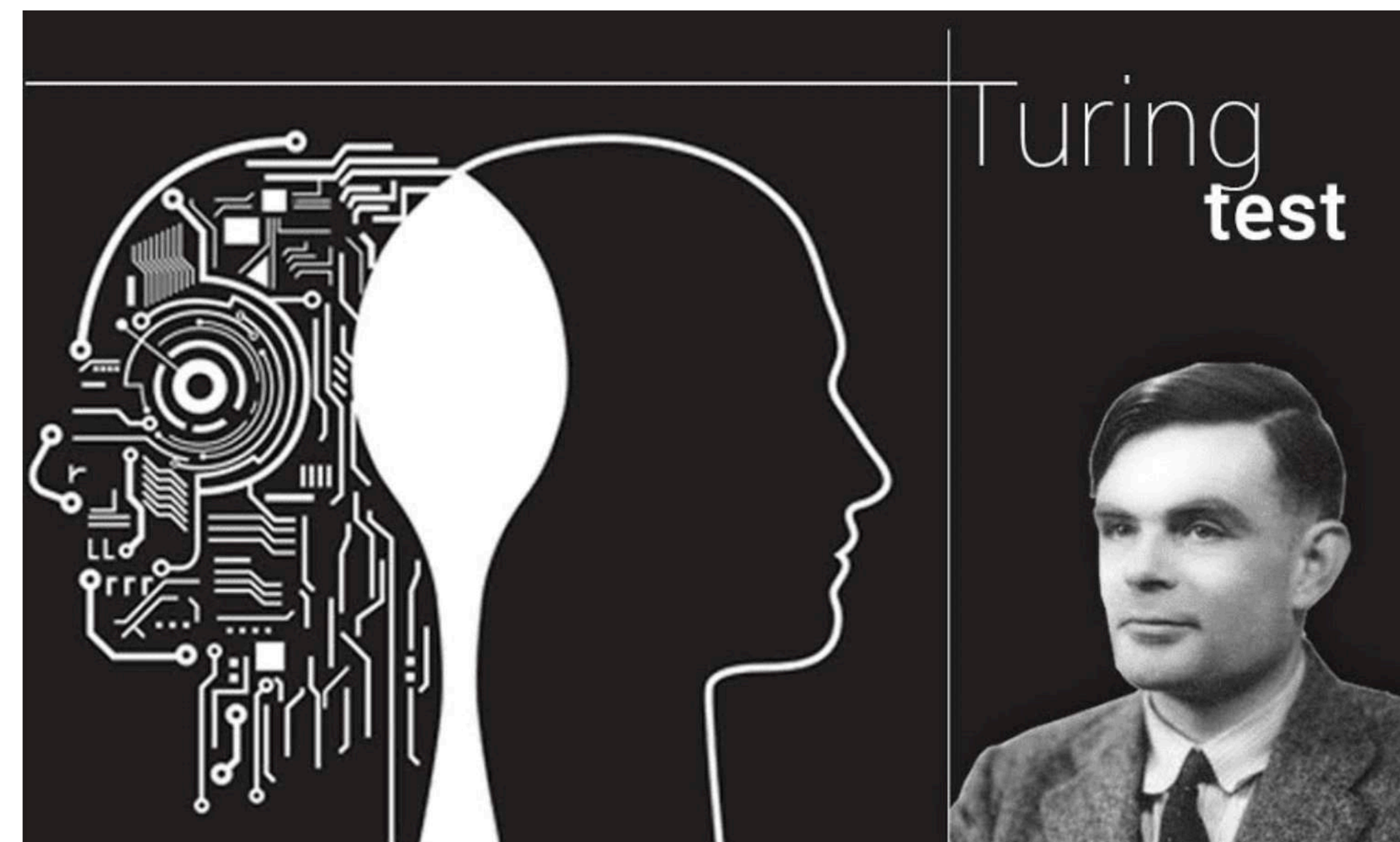


* This Presentation is mainly drawn from the research paper named ” *An Introduction to Ethics in Robotics and AI* “ by Bartneck, Wagner, Welsh, Lütge.

Introduction

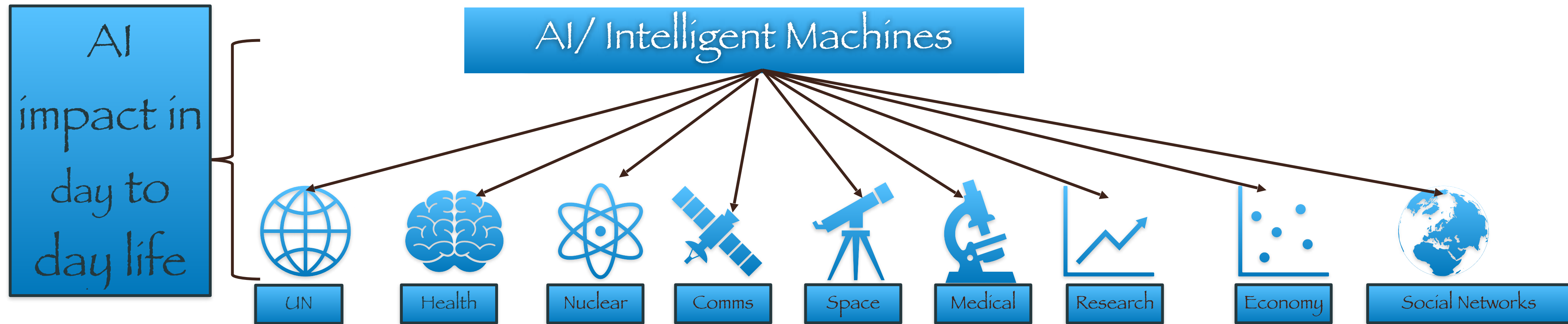
1. Intelligent Machine?

2. The Turing test



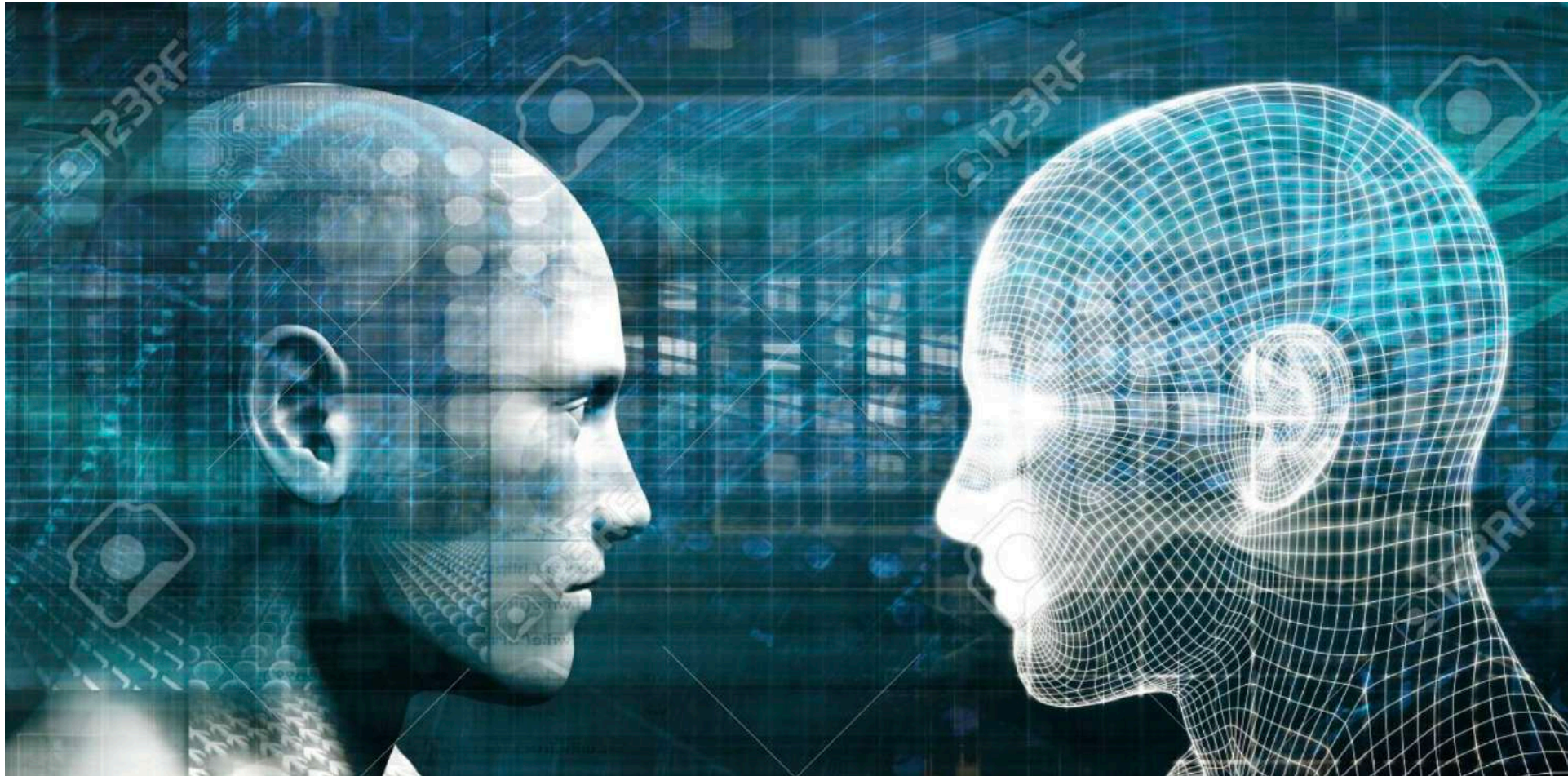
3. AI is everywhere

AI – the new wheel of 21st Century.



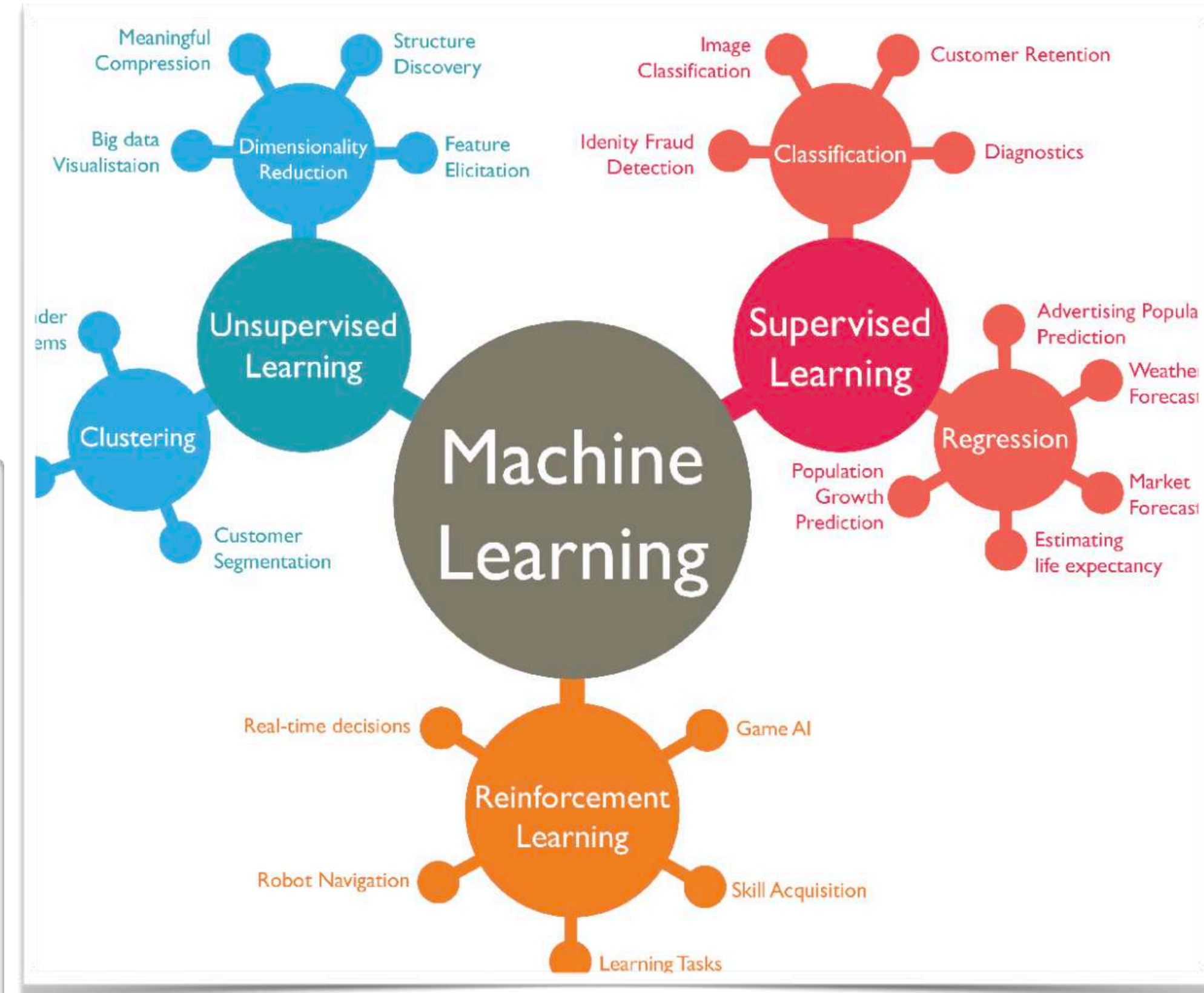
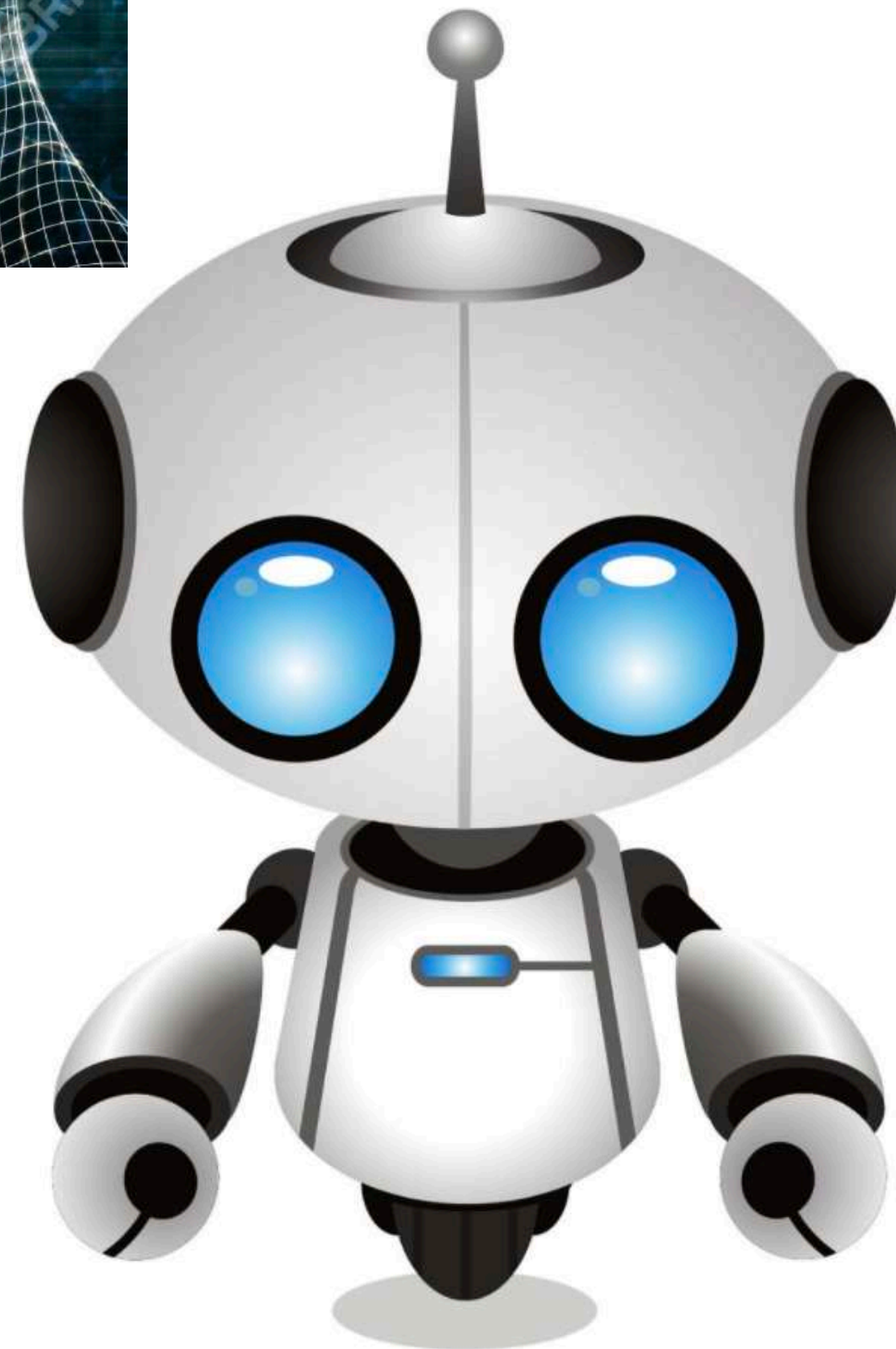
- The **penetration** of Intelligent machine /AI in today's world impacts in all walks of our daily lives.
- **For instance**, we carry AI in our pocket in the form of smart devices.
- Allows customer to **personalise** their worlds.

Introduction



4. Strong and Weak AI

5. Robot



6. Machine Learning

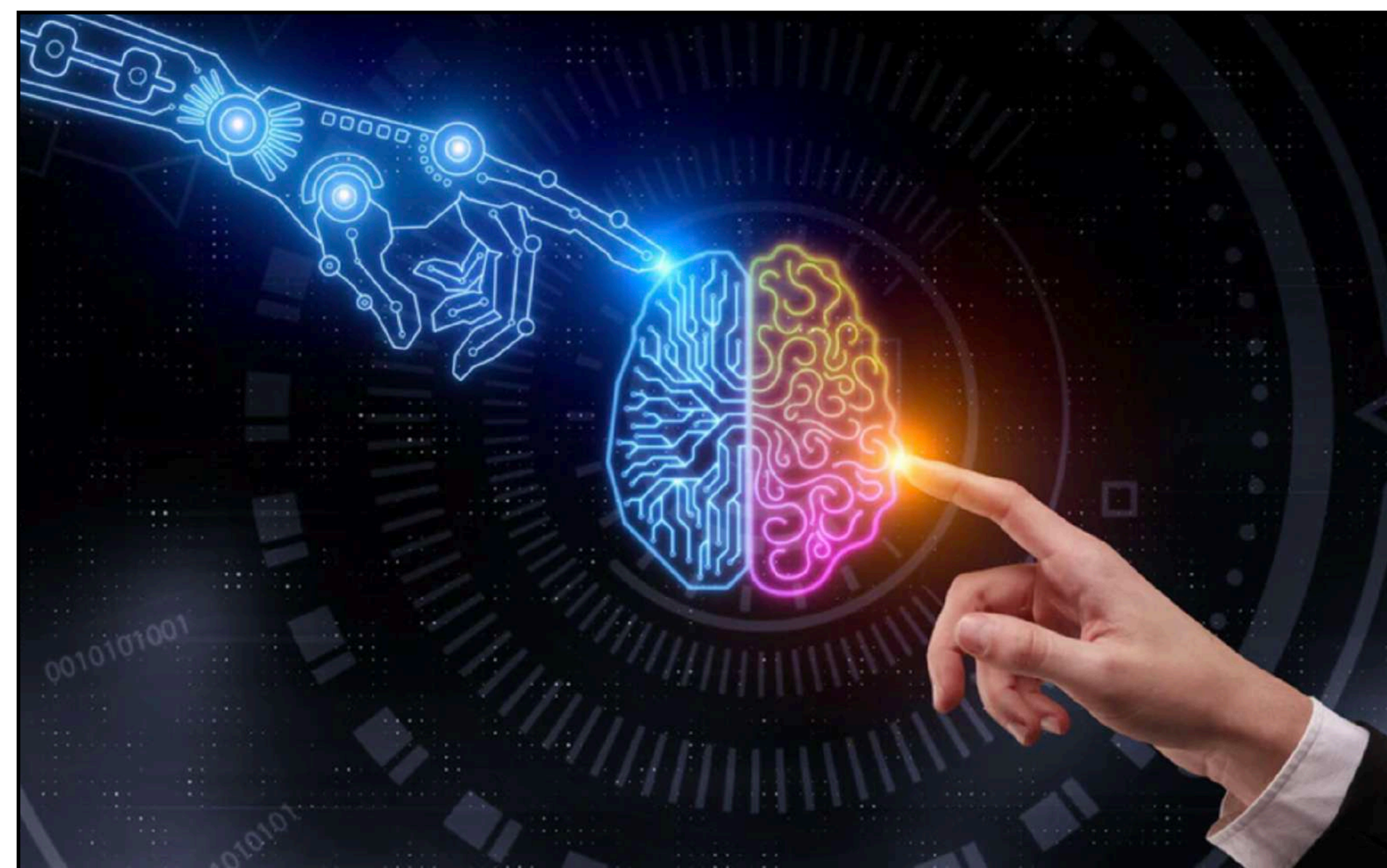
Machine Learning

- The ability to automatically **learn and improve** from experience without being explicitly programmed.
- Focuses on the **development** of computer programs that can access data and use it **learn** for themselves.
- Sub-categorized into **three** types of learning:
 - Supervised learning
 - Unsupervised learning
 - Reinforcement learning



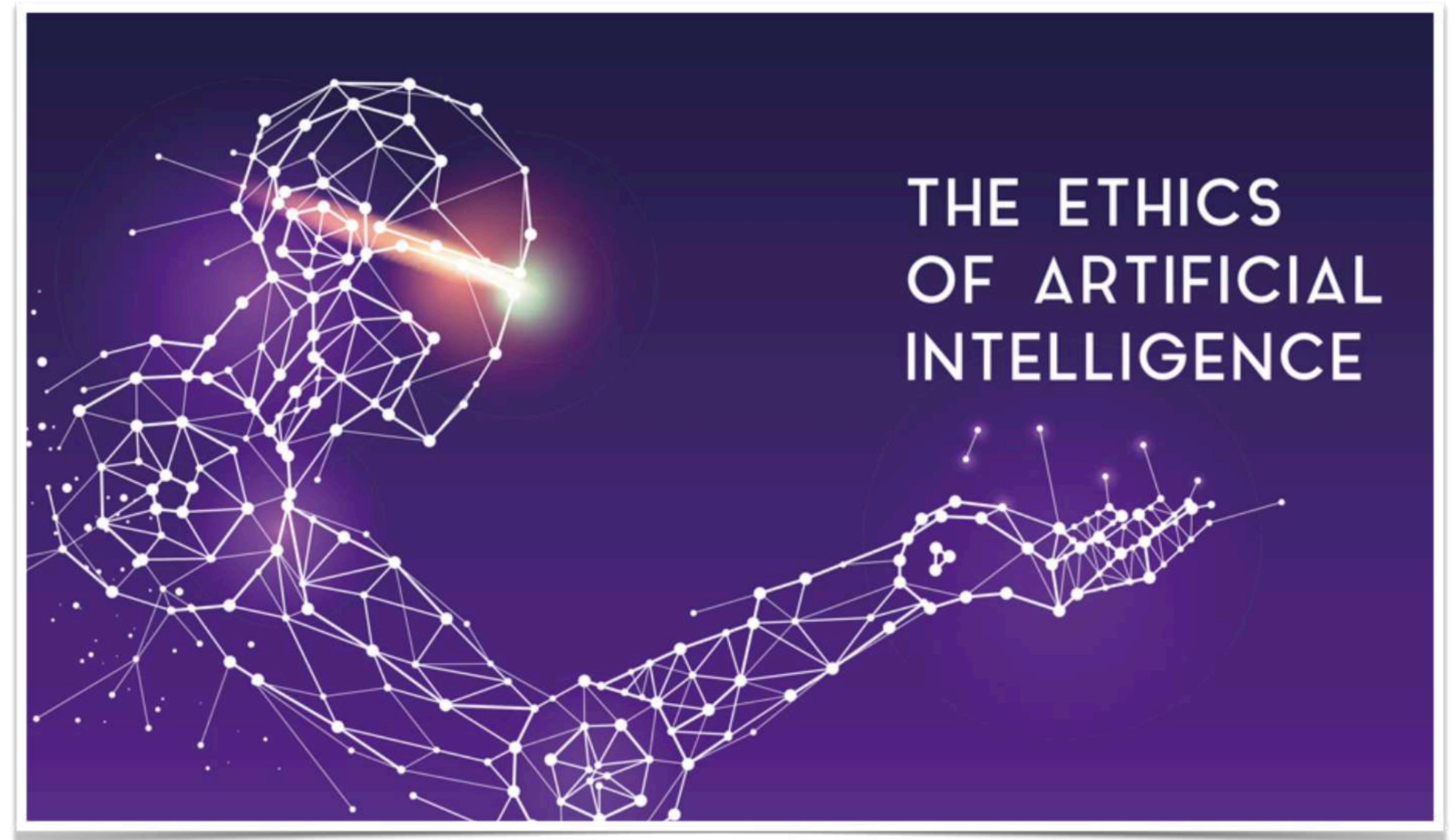
Ethics in AI.

- What is Ethics?
 - The discipline dealing with right vs wrong.
 - Ethics” and “Morality”.
 - Complex set of rules, values and norms.



Types of Ethics

- Descriptive Ethics
- Normative Ethics
- Meta-ethics
- Applied Ethics
- Ethics and Law
- **Machine Ethics**



Machine Ethics

- Adding **moral** behaviours to **machines**.
- How to **design** an **AI** that is **capable** of making **moral decisions**?
 - Machine Ethics Examples
 - Moral Diversity and Testing



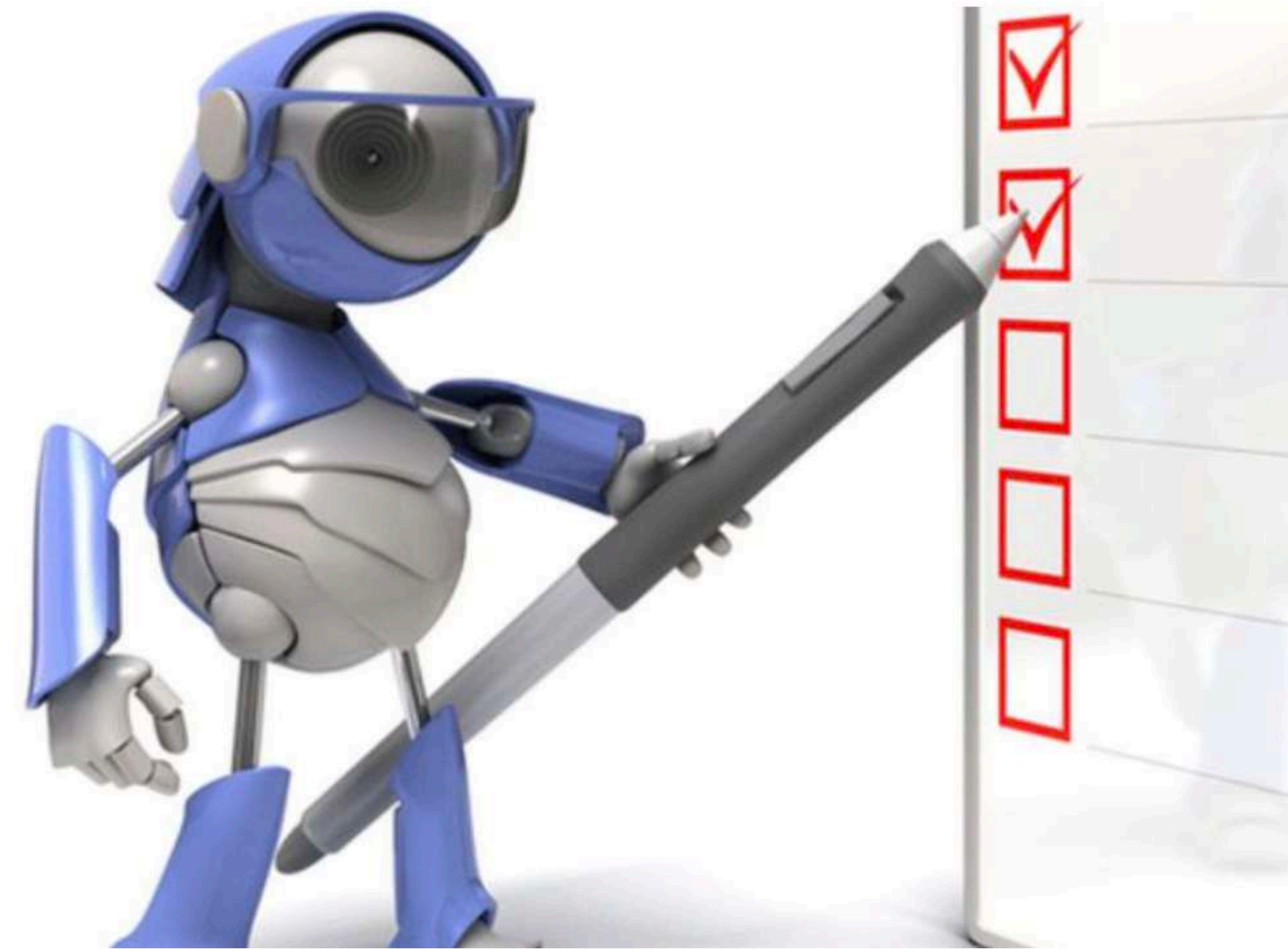
Why machine Ethics?

- **Trust and Fairness in AI Systems**
 - Help achieve an individual's goals in a situation by uncertainty and vulnerability.
 - Not certain that a system will act in our best interest.
 - Need of ethics in Machines
 - Critical to user acceptance.
 - Not compromise their autonomy.



Why machine Ethics?..

- **Five Ethical Principles for Trustworthy and Fair AI.**
 - Non-maleficence
 - Beneficence
 - Autonomy
 - Justice
 - Explicability



<https://www.bbc.com/news/technology-32334568>

Why machine Ethics?..

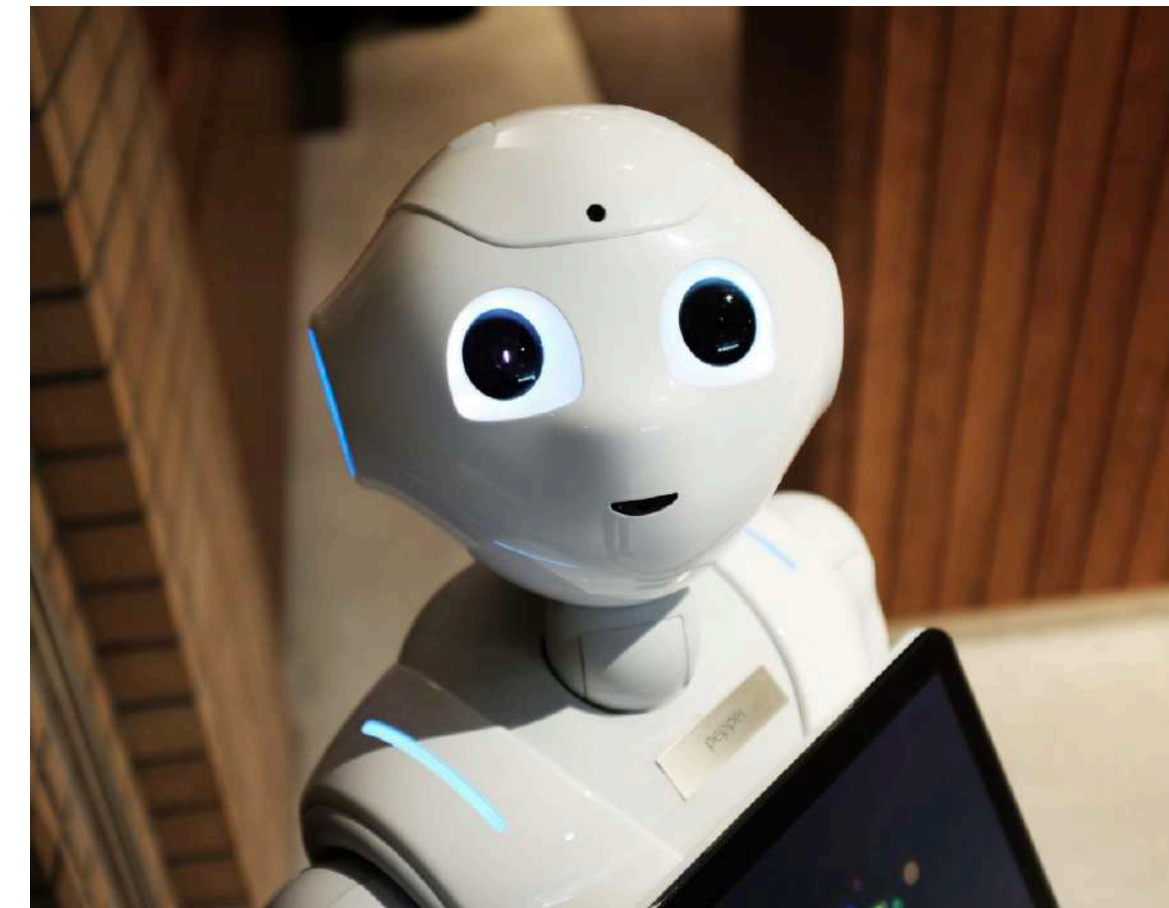
- **Responsibility and Liability**
 - Not concerns with liability when everything is goes fine.
 - Comes into play only when something goes wrong.
 - So what does “**go wrong**” mean in the case of AI technologies?



Why machine Ethics?..

- **Handling Privacy Issues**

- Right to not be observed.
- People often act in a manner as to increase their own privacy.
- Privacy is valuable for a number of important reasons.
- Allows people to make their own, non-coerced decisions.
- To better calculate behaviour.
- To take decisions and actions that do not conform to certain social norms.



Why machine Ethics?..

- **Handling Privacy Issues ..**

- Gathering personal data become dramatically easier.
- Possible to track every step users take and every restaurant they visit.
- Users of social networks often transfer the copyright of data to the platform provider.
- Example: The Chinese government started work on a Social Credit System in 2014



Role of Ethics

- Development of AI in full respect of fundamental rights.
- Global ethical standard
- Lack of practical, agreed guidelines and rules regarding systems.
- Regulatory aspects of AI just started.
- Industry requires clear rules for speedy innovation based on AI.
- Companies may steer away from AI applications.
- Ethics becomes important at many layers of the policy discussion.



Role of Ethics..

- A topic for the engineer designing the system.
- Value the impacts of AI technology.
- Ethics in AI is much broader and concerns very basic design choices.



Other issues

- Existing AI community definition of what constitutes ethics or morality is not clear.
- Definitions ranges from one extreme to another with some references.
- Leaves room for different distribution of the benefits.
- How to negotiate multiple and possible conflicting ideals, values, and notions of what is good.



Conclusion

- No clear task specification for moral behaviour.
- Treat ethics as hard constraints, never to be violated, or soft constraints.
- Computationally intractable, putting ethical principles beyond the limits of effective computation.



Conclusion

- Should explore the strengths and weaknesses of the various approaches to programming AMAs.
- Mental faculties difficult to simulate.
- Essential to cultivate the right feelings, sentiments, and virtues.
- To develop criteria and tests for evaluating an artificial entity's moral aptitude.
- Scientific knowledge about human cognitive and emotional faculties has grown exponentially.
- Designing artificial systems that function convincingly and autonomously in real physical and social environments are complex and difficult.

Summary

- Integrate artificial moral agents into new technologies.
- AMAs should be able to make decisions that honour privacy, uphold shared ethical standards, protect civil rights and individual liberty.
- Help engineers become aware of their work's ethical dimensions.
- Sensitive to systems' choices.

Summary..

- Self-governing machines, capable of assessing the ethical acceptability of the options they face.
- Focus toward more human intervention in computers and robots' decision making processes.
- Extending moral agency to artificial entities raises many new issues.
- For example, what are appropriate criteria for determining success in creating an AMA?

References

- “*An Introduction to Ethics in Robotics and AI*” by Bartneck, C., Lütge, C., Wagner, A., Welsh, S. **Url:** <https://www.bartneck.de/publications/2020/An-Introduction-To-Ethics-In-Robotics-And-AI/An-Introduction-To-Ethics-In-Robotics-And-AI.pdf>
- <https://www.semanticscholar.org/paper/Why-Machine-Ethics-Allen-Wallach/6609ffd2e42fa9552b613d5ceeaee39ee8515b9a>
- [https://en.wikipedia.org/wiki/Machine_ethics#:~:text=Machine%20ethics%20\(or%20machine%20morality,known%20as%20artificial%20intelligent%20agents](https://en.wikipedia.org/wiki/Machine_ethics#:~:text=Machine%20ethics%20(or%20machine%20morality,known%20as%20artificial%20intelligent%20agents).
- <https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0>
- And General Google research