**Mapping Core Values to Ethical Considerations for AI, Incorporating Social Values, Cultural Differences, Personal Values, New Directions in Digital Ethics, General Cognitive Abilities, Genuine Reasoning, Artificial Superintelligence, and now, Marvin Minsky's Framework**

# Trust

## Normative:

* + **Beliefs:** AI systems should be built on trust, ensuring transparency and reliability.
  + **Values:** Accuracy, Reliability, Safety, Predictability, **Security**, **Resilience**, **Confidentiality**, **Integrity**, **Transparency**, **Non-repudiation**.
  + **Principles:** The precautionary principle should guide AI development, prioritizing safety and minimizing risk.
  + **Ideals:** An ideal AI system would be perfectly safe, reliable, and trustworthy.
  + **Localized Autonomy:** Communities should have a say in how AI systems are developed and deployed within their context.
  + **Social Values:** **Trust** (Relational), **Transparency** (Structural), **Accountability** (Structural), **Dignity** (Structural), **Solidarity** (Societal)
  + **Cultural Differences:** Trust-building may vary across cultures. Western cultures might prioritize individual expertise, while Eastern cultures might emphasize collective trust and harmonious relationships.
  + **Personal Values:** Individual values related to honesty, integrity, and trustworthiness should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to build trust in the context of a global network society, considering issues like algorithmic accountability, data sovereignty, and the role of diverse stakeholders.
  + **General Cognitive Abilities:**
    - **Reasoning & Problem Solving:** AI should be able to demonstrate logical reasoning and problem-solving abilities to build trust in its decision-making.
    - **Knowledge Acquisition:** AI should be able to learn and adapt, demonstrating a grasp of the complexities of trust and its nuances in different contexts.
    - **Metacognition:** AI should be able to self-assess its own performance and identify areas where trust might be eroded.
    - **Adaptive Behavior:** AI should be able to adapt its behavior based on user feedback and changing trust levels.
    - **Abstract Thinking:** AI should be able to understand and navigate the abstract concept of trust and its implications for human-AI interaction.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning, considering the ethical implications of its actions and decisions on trust. This includes avoiding manipulation, deception, or exploitation.
  + **Artificial Superintelligence:** In the context of ASI, trust becomes paramount. Mechanisms for ensuring transparency, control, and alignment with human values are crucial to maintain trust and prevent unintended consequences.
  + **Marvin Minsky Framework:** Trust in ASI may require understanding the internal workings of its "society of mind," ensuring that individual agents within the system operate ethically and transparently.

## Regulatory:

* + **Laws:** Laws should mandate safety standards, risk assessments, and liability frameworks for AI developers.
  + **Rules:** Clear rules should govern the use of AI in critical sectors like healthcare and transportation.
  + **Standards:** Technical standards should be established for AI safety, reliability, explainability, **security**, and **resilience**.
  + **Compliance:** Strict compliance with regulations and standards is crucial to ensure public trust.
  + **Entitlements:** Citizens should have the right to information about how AI systems are used and the right to challenge decisions made by AI systems.
  + **Freedoms:** Individuals should have the freedom from harm caused by AI systems.
  + **Ethical Localization:** Regulations should consider the unique cultural and social contexts of different regions.

## Behavioral:

* + **Ethics:** Developers should adhere to a strong ethical code of conduct, prioritizing safety and public interest.
  + **Morals:** Moral considerations should guide the development and deployment of AI systems, avoiding harm and promoting well-being.
  + **Virtues:** Virtues such as honesty, integrity, **empathy**, and **responsibility** are crucial for ethical AI development.
  + **Norms:** Industry norms and best practices should promote responsible AI development and deployment.
  + **Customs:** Cultural and social norms should be considered when developing and deploying AI systems.

## Conceptual:

* + **Foundational Concepts (Rights, Responsibilities):** Rights (e.g., right to privacy, right to be free from discrimination), Responsibilities (e.g., responsibility of developers to ensure safety, reliability, **security**, and **resilience**).
  + **Operational Concepts (Obligations, Privileges):** Obligations (e.g., obligation to mitigate bias, obligation to ensure transparency, **security**, and **resilience**), Privileges (e.g., privileged access to data for research purposes, subject to ethical guidelines).
  + **Contextual and Evaluative Concepts (Consequences, Intentions, Duties, Expectations, Practices, Directives):** Consequences (e.g., potential societal impacts, environmental impacts), Intentions (e.g., the intended purpose of the AI system), Duties (e.g., duty of care, duty to inform), Expectations (e.g., public expectations regarding AI safety, reliability, and **security**), Practices (e.g., best practices for AI development and deployment), Directives (e.g., ethical guidelines, codes of conduct).
  + **AI Ethics:** Ensure that AI development aligns with ethical principles, minimizing unintended consequences and promoting societal benefit.
  + **Cultural Differences:** Consider the cultural context when evaluating trust. In some cultures, trust may be built through long-term relationships and shared experiences, while in others, it may be based on individual expertise and credentials.
  + **Personal Values:** Individual values related to honesty, integrity, and trustworthiness should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to build trust in the context of a global network society, considering issues like algorithmic accountability, data sovereignty, and the role of diverse stakeholders.
  + **General Cognitive Abilities:** Reasoning & Problem Solving, Knowledge Acquisition, Metacognition, Adaptive Behavior, Abstract Thinking are crucial for AI to demonstrate trustworthiness and build human trust.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning, considering the ethical implications of its actions and decisions on trust. This includes avoiding manipulation, deception, or exploitation.
  + **Artificial Superintelligence:** In the context of ASI, trust becomes paramount. Mechanisms for ensuring transparency, control, and alignment with human values are crucial to maintain trust and prevent unintended consequences.
  + **Marvin Minsky Framework:** Understanding how an ASI's "society of mind" operates is crucial for building trust. Individual agents within the system must be ethical and transparent in their decision-making, and the overall system must be designed to avoid unintended consequences and maintain trust with humans.

# Fairness

## Normative:

* **Beliefs:** AI systems should not discriminate against or unfairly disadvantage any individual or group.
* **Values:** Equity, Justice, Inclusion, **Fairness**, **Bias Mitigation**.
* **Principles:** The principle of equal opportunity should guide the development and deployment of AI systems.
* **Ideals:** An ideal AI system would promote social justice and eliminate existing inequalities.
* **Localized Autonomy:** Communities should have a say in how AI systems are used within their context, ensuring that local values and needs are considered.
* **Social Values:** **Fairness** (Relational), **Justice** (Structural), **Equity** (Structural), **Equality** (Relational), **Dignity** (Structural)
* **Cultural Differences:** Concepts of fairness and justice may vary across cultures. Some cultures may prioritize group needs over individual rights, while others may emphasize individual autonomy and equal opportunity.
* **Personal Values:** Individual values related to fairness, justice, and equality should guide personal interactions with AI systems.
* **New Directions in Digital Ethics:** Explore how to ensure fairness and equity in a globalized and interconnected world, considering issues like algorithmic bias, data colonialism, and the digital divide.
* **General Cognitive Abilities:**
  + **Reasoning & Problem Solving:** AI should be able to identify and address biases in data and algorithms to ensure fair and equitable outcomes.
  + **Knowledge Acquisition:** AI should be able to learn and adapt to changing societal norms and values related to fairness and justice.
  + **Metacognition:** AI should be able to self-assess for biases in its own decision-making processes.
  + **Adaptive Behavior:** AI should be able to adjust its behavior to ensure fair and equitable outcomes for all users.
  + **Abstract Thinking:** AI should be able to understand and navigate the complex social and ethical issues related to fairness and justice.
* **Genuine Reasoning:** AI should be able to engage in genuine reasoning to identify and address potential biases in its decision-making, ensuring fair and equitable outcomes for all individuals and groups.
* **Artificial Superintelligence:** In the context of ASI, ensuring fairness and equity is paramount. It is crucial to prevent the emergence of biases that could lead to discrimination, oppression, or the exacerbation of existing inequalities. The concept of "fairness" for an ASI may require a deeper understanding of societal values and the potential impact of AI on social structures.

## Regulatory:

* + **Laws:** Laws should prohibit discrimination based on protected characteristics and require measures to mitigate bias in AI systems.
  + **Rules:** Clear rules should govern the use of AI in sensitive areas like hiring and lending, ensuring fairness and preventing discrimination.
  + **Standards:** Standards for fairness and bias mitigation should be developed and implemented across the AI ecosystem.
  + **Compliance:** Compliance with anti-discrimination laws and regulations is essential.
  + **Entitlements:** Individuals should have the right to challenge discriminatory decisions made by AI systems.
  + **Freedoms:** Individuals should be free from discrimination based on factors like race, gender, and religion.
  + **Ethical Localization:** Regulations should consider the unique social and cultural contexts of different regions and communities.

## Behavioral:

* + **Ethics:** Developers should actively work to identify and mitigate biases in data and algorithms.
  + **Morals:** A strong moral compass is necessary to avoid perpetuating existing societal biases through AI.
  + **Virtues:** Virtues like empathy, compassion, and a commitment to social justice are crucial for ethical AI development.
  + **Norms:** Industry norms should promote the development and use of fair and unbiased AI systems.
  + **Customs:** Cultural and social norms should be considered to ensure that AI systems are culturally sensitive and do not perpetuate harmful stereotypes.

## Conceptual:

* + **Foundational Concepts (Rights, Responsibilities):** Rights (e.g., right to equal protection under the law), Responsibilities (e.g., responsibility to avoid perpetuating or amplifying existing biases).
  + **Operational Concepts (Obligations, Privileges):** Obligations (e.g., obligation to mitigate bias, obligation to ensure equitable access), Privileges (e.g., the privilege to use AI for positive social change).
  + **Contextual and Evaluative Concepts (Consequences, Intentions, Duties, Expectations, Practices, Directives):** Consequences (e.g., the potential for AI to exacerbate existing inequalities), Intentions (e.g., the intended purpose of the AI system and its potential impact on different groups), Duties (e.g., duty to ensure fairness and avoid discrimination), Expectations (e.g., public expectations regarding the ethical use of AI), Practices (e.g., best practices for bias detection and mitigation), Directives (e.g., ethical guidelines for fair AI development).
  + **AI Ethics:** Ensure that AI systems are developed and deployed in a way that is fair and equitable across different cultural contexts.
  + **Cultural Differences:** Consider how cultural values and norms may influence perceptions of fairness and justice. Some cultures may prioritize group needs over individual rights, while others may emphasize individual autonomy and equal opportunity.
  + **Personal Values:** Individual values related to fairness, justice, and equality should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to ensure fairness and equity in a globalized and interconnected world, considering issues like algorithmic bias, data colonialism, and the digital divide.
  + **General Cognitive Abilities:** Reasoning & Problem Solving, Knowledge Acquisition, Metacognition, Adaptive Behavior, Abstract Thinking are crucial for AI to identify and address biases.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning to identify and address potential biases in its decision-making, ensuring fair and equitable outcomes for all individuals and groups.
  + **Artificial Superintelligence:** In the context of ASI, ensuring fairness and equity is paramount. It is crucial to prevent the emergence of biases that could lead to discrimination, oppression, or the exacerbation of existing inequalities. The concept of "fairness" for an ASI may require a deeper understanding of societal values and the potential impact of AI on social structures.
  + **Marvin Minsky Framework:** Minsky's concept of multiple interacting agents within an AI system highlights the importance of ensuring fairness and equity at the level of individual agents and their interactions. Biases within individual agents can propagate through the system and lead to unfair or discriminatory outcomes. Therefore, it is crucial to ensure that each agent within an ASI operates in a fair and equitable manner, and that the interactions between agents do not result in systemic biases.

# Transparency

## Normative:

* + **Beliefs:** Transparency is crucial for building trust and understanding how AI systems work.
  + **Values:** Openness, Honesty, **Transparency**, **Accountability**.
  + **Principles:** The principle of transparency should guide the development and deployment of AI systems.
  + **Ideals:** An ideal AI system would be completely transparent, allowing users to fully understand its decision-making process.
  + **Localized Autonomy:** Communities should have access to information about how AI systems are used within their context.
  + **Social Values:** **Transparency** (Structural), **Accountability** (Structural), **Dignity** (Structural), **Rule of Law** (Structural)
  + **Cultural Differences:** Transparency may be interpreted differently across cultures. Some cultures may prioritize open communication and information sharing, while others may place greater emphasis on maintaining social harmony and avoiding conflict.
  + **Personal Values:** Individual values related to openness, honesty, and transparency should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to achieve transparency in complex AI systems while protecting sensitive information and intellectual property.
  + **General Cognitive Abilities:**
    - **Reasoning & Problem Solving:** AI should be able to explain its reasoning and decision-making processes in a clear and understandable manner.
    - **Knowledge Acquisition:** AI should be able to learn and adapt to different levels of transparency required by users and stakeholders.
    - **Metacognition:** AI should be able to reflect on its own decision-making processes and identify areas where transparency can be improved.
    - **Adaptive Behavior:** AI should be able to adjust its level of transparency based on user needs and preferences.
    - **Abstract Thinking:** AI should be able to understand and navigate the complex issues surrounding transparency, including privacy, security, and the potential for misuse of information.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about the importance of transparency and its role in building trust and accountability. This includes considering the potential benefits and risks of transparency in different contexts.
  + **Artificial Superintelligence:** In the context of ASI, transparency becomes even more critical. Mechanisms for understanding and explaining the reasoning and decision-making processes of an ASI are essential for maintaining trust and ensuring that its actions are aligned with human values. Minsky's framework of multiple interacting agents suggests that transparency may require understanding the reasoning and decision-making processes of individual agents within the system, as well as the overall system behavior.

## Regulatory:

* + **Laws:** Laws should require transparency about the data used to train AI systems, the algorithms that govern their decision-making, and the rationale behind their outputs (to the extent possible).
  + **Rules:** Clear rules should govern the disclosure of information about AI systems to the public.
  + **Standards:** Standards for transparency and explainability should be developed and implemented.
  + **Compliance:** Compliance with transparency regulations is essential.
  + **Entitlements:** Individuals should have the right to access information about how AI systems are used and how decisions affecting them are made.
  + **Freedoms:** Individuals should have the freedom to access information about AI systems and their potential impacts.
  + **Ethical Localization:** Transparency requirements should consider the unique cultural and social contexts of different regions.
* **Behavioral:**
  + **Ethics:** Developers should prioritize transparency and be willing to communicate openly about AI systems and their limitations.
  + **Morals:** A commitment to transparency is a moral imperative for ethical AI development.
  + **Virtues:** Virtues like honesty, openness, and a willingness to engage in public discourse are crucial.
  + **Norms:** Industry norms should promote transparency and open communication about AI systems.
  + **Customs:** Cultural norms regarding openness and information sharing should be considered.

## Conceptual:

* + **Foundational Concepts (Rights, Responsibilities):** Rights (e.g., right to information, right to know how decisions affecting you are made), Responsibilities (e.g., responsibility to provide transparent information about AI systems).
  + **Operational Concepts (Obligations, Privileges):** Obligations (e.g., obligation to explain AI systems to the extent possible), Privileges (e.g., privileged access to data for research purposes, subject to appropriate transparency measures).
  + **Contextual and Evaluative Concepts (Consequences, Intentions, Duties, Expectations, Practices, Directives):** Consequences (e.g., the potential impact of transparency on innovation and competition), Intentions (e.g., the intended level of transparency and the reasons for it), Duties (e.g., duty to explain, duty to inform), Expectations (e.g., public expectations regarding transparency in AI), Practices (e.g., best practices for communicating complex technical information to the public), Directives (e.g., ethical guidelines for transparency in AI).
  + **AI Ethics:** Ensure that AI development aligns with ethical principles, minimizing unintended consequences and promoting societal benefit.
  + **Cultural Differences:** Consider the cultural context when evaluating trust. In some cultures, trust may be built through long-term relationships and shared experiences, while in others, it may be based on individual expertise and credentials.
  + **Personal Values:** Individual values related to honesty, openness, and transparency should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to achieve transparency in complex AI systems while protecting sensitive information and intellectual property.
  + **General Cognitive Abilities:** Reasoning & Problem Solving, Knowledge Acquisition, Metacognition, Adaptive Behavior, Abstract Thinking are crucial for AI to explain its reasoning and decision-making processes in a clear and understandable manner.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about the importance of transparency and its role in building trust and accountability. This includes considering the potential benefits and risks of transparency in different contexts.
  + **Artificial Superintelligence:** In the context of ASI, transparency becomes even more critical. Mechanisms for understanding and explaining the reasoning and decision-making processes of an ASI are essential for maintaining trust and ensuring that its actions are aligned with human values. Minsky's framework of multiple interacting agents suggests that transparency may require understanding the reasoning and decision-making processes of individual agents within the system, as well as the overall system behavior.

# Accountability

## Normative:

* + **Beliefs:** There must be clear lines of responsibility for the development, deployment, and use of AI systems.
  + **Values:** Accountability, Responsibility, Justice, **Auditability**, **Non-repudiation**.
  + **Principles:** The principle of accountability should guide the development and deployment of AI systems.
  + **Ideals:** An ideal AI system would be fully accountable for its actions and decisions.
  + **Localized Autonomy:** Communities should have a voice in how AI systems are used within their context and should have recourse in case of harm.
  + **Social Values:** **Accountability** (Structural), **Justice** (Structural), **Rule of Law** (Structural), **Dignity** (Structural)
  + **Cultural Differences:** Concepts of accountability and responsibility may vary across cultures. Some cultures may emphasize individual responsibility, while others may prioritize collective responsibility.
  + **Personal Values:** Individual values related to responsibility, accountability, and ownership of one's actions are crucial for ethical interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to establish accountability in complex AI systems, including the development of effective auditing and oversight mechanisms.
  + **General Cognitive Abilities:**
    - **Reasoning & Problem Solving:** AI should be able to analyze its own actions and decisions, identifying potential risks and areas for improvement.
    - **Knowledge Acquisition:** AI should be able to learn from its mistakes and adapt its behavior accordingly.
    - **Metacognition:** AI should be able to self-assess its own performance and identify areas where accountability is lacking.
    - **Adaptive Behavior:** AI should be able to adjust its behavior to ensure accountability for its actions and decisions.
    - **Abstract Thinking:** AI should be able to understand and navigate the complex ethical and legal issues related to accountability.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about its own actions and decisions, taking responsibility for its outcomes and identifying potential areas for improvement.
  + **Artificial Superintelligence:** In the context of ASI, accountability becomes even more critical. Mechanisms for ensuring that ASI systems are aligned with human values and remain accountable for their actions are essential. This may involve developing sophisticated auditing and control systems, as well as establishing clear lines of responsibility for the development, deployment, and oversight of ASI systems. Minsky's framework of multiple interacting agents suggests that accountability may be distributed across different levels of the system, with individual agents responsible for their own actions and the system as a whole accountable for its overall behavior.

## Regulatory:

* + **Laws:** Laws should establish clear accountability mechanisms, including the ability to identify and address harms caused by AI systems.
  + **Rules:** Clear rules should govern the development, deployment, and use of AI systems, ensuring that accountability is built into all stages of the AI lifecycle.
  + **Standards:** Standards for AI safety, reliability, and accountability should be developed and implemented.
  + **Compliance:** Compliance with accountability regulations is essential.
  + **Entitlements:** Individuals should have the right to seek redress for harms caused by AI systems and to hold developers and deployers accountable.
  + **Freedoms:** Individuals should have the freedom from harm caused by unaccountable AI systems.
  + **Ethical Localization:** Accountability mechanisms should consider the unique legal and regulatory frameworks of different regions.

## Behavioral:

* + **Ethics:** Developers should accept responsibility for the potential impacts of their AI systems and strive to mitigate any potential harms.
  + **Morals:** A strong moral compass is necessary to ensure that developers and deployers of AI systems act responsibly and ethically.
  + **Virtues:** Virtues like responsibility, accountability, and a commitment to ethical conduct are crucial for all stakeholders involved in AI development and deployment.
  + **Norms:** Industry norms should promote responsible AI development and deployment, including mechanisms for accountability and oversight.
  + **Customs:** Cultural norms related to responsibility and accountability should be considered when developing and deploying AI systems.

## Conceptual:

* + **Foundational Concepts (Rights, Responsibilities):** Rights (e.g., right to redress for harms caused by AI systems), Responsibilities (e.g., responsibility of developers to ensure the safety and reliability of their AI systems, responsibility of users to use AI systems ethically).
  + **Operational Concepts (Obligations, Privileges):** Obligations (e.g., obligation to mitigate risks, obligation to address harms caused by AI systems), Privileges (e.g., the privilege to use AI for beneficial purposes, subject to appropriate accountability measures).
  + **Contextual and Evaluative Concepts (Consequences, Intentions, Duties, Expectations, Practices, Directives):** Consequences (e.g., the potential societal and ethical consequences of AI systems), Intentions (e.g., the intended use of AI systems and their potential impact on society), Duties (e.g., duty to ensure the safe and responsible development and deployment of AI systems), Expectations (e.g., public expectations regarding the accountability of AI developers and deployers), Practices (e.g., best practices for AI risk assessment and management, mechanisms for accountability and redress), Directives (e.g., ethical guidelines for AI development and deployment, codes of conduct).
  + **AI Ethics:** Ensure that AI development and deployment are conducted in a responsible and accountable manner, with clear lines of responsibility for all stakeholders.
  + **Cultural Differences:** Consider how cultural norms and values related to responsibility and accountability may influence the development and deployment of AI systems.
  + **Personal Values:** Individual values related to responsibility, accountability, and ownership of one's actions should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore how to establish accountability in complex AI systems, including the development of effective auditing and oversight mechanisms.
  + **General Cognitive Abilities:** Reasoning & Problem Solving, Knowledge Acquisition, Metacognition, Adaptive Behavior, Abstract Thinking are crucial for AI to understand and navigate the complexities of accountability.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about its own actions and decisions, taking responsibility for its outcomes and identifying potential areas for improvement.
  + **Artificial Superintelligence:** In the context of ASI, accountability becomes even more critical. Mechanisms for ensuring that ASI systems are aligned with human values and remain accountable for their actions are essential. This may involve developing sophisticated auditing and control systems, as well as establishing clear lines of responsibility for the development, deployment, and oversight of ASI systems. Minsky's framework of multiple interacting agents suggests that accountability may be distributed across different levels of the system, with individual agents responsible for their own actions and the system as a whole accountable for its overall behavior.

# Non-Maleficence

* **Normative:**
  + **Beliefs:** AI systems should be developed and deployed in a way that minimizes harm to humans and the environment.
  + **Values:** Safety, Beneficence, Non-maleficence, Well-being, Sustainability.
  + **Principles:** The principle of non-maleficence, the ethical obligation to do no harm, should be a central guiding principle in AI development.
  + **Ideals:** An ideal AI system would not only be beneficial but also completely harmless.
  + **Localized Autonomy:** Communities should have a say in how AI systems are developed and deployed within their context, ensuring that local values and concerns about potential harm are addressed.
  + **Social Values:** **Dignity** (Structural), **Solidarity** (Societal), **Sustainability** (Aspirational), **Peace** (Aspirational).
  + **Cultural Differences:** Perceptions of harm and well-being can vary across cultures. It is crucial to consider these differences when developing and deploying AI systems.
  + **Personal Values:** Individual values related to compassion, empathy, and the avoidance of harm should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore the long-term societal and environmental impacts of AI, and develop strategies to mitigate potential harms.
  + **General Cognitive Abilities:**
    - **Reasoning & Problem Solving:** AI should be able to anticipate and mitigate potential risks and harms associated with its actions and decisions.
    - **Knowledge Acquisition:** AI should be able to learn and adapt to changing understandings of harm and risk.
    - **Metacognition:** AI should be able to self-assess its own potential for causing harm and take steps to minimize or avoid it.
    - **Adaptive Behavior:** AI should be able to adjust its behavior to minimize harm and maximize benefits.
    - **Abstract Thinking:** AI should be able to understand and navigate the complex ethical issues related to harm and well-being.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about the potential harms of its actions and decisions, and strive to act in ways that minimize or avoid harm.
  + **Artificial Superintelligence:** The principle of non-maleficence takes on even greater significance in the context of ASI. It is crucial to ensure that ASI systems are developed and deployed in a way that minimizes the risk of existential harm to humanity and the environment. This may involve developing robust safety mechanisms, ensuring alignment with human values, and carefully considering the long-term consequences of ASI development. Minsky's framework of multiple interacting agents suggests that ensuring non-maleficence may require careful coordination and control of individual agents within the system to prevent unintended consequences and avoid cascading failures.
* **Regulatory:**
  + **Laws:** Laws should mandate safety standards and risk assessments to minimize the potential for harm caused by AI systems.
  + **Rules:** Clear rules should govern the development and deployment of AI systems in high-risk domains, such as autonomous weapons and critical infrastructure.
  + **Standards:** Safety and risk management standards should be developed and implemented across the AI ecosystem.
  + **Compliance:** Compliance with safety and risk management regulations is essential.
  + **Entitlements:** Individuals should have the right to be protected from harm caused by AI systems.
  + **Freedoms:** Individuals should have the freedom from harm caused by AI systems.
  + **Ethical Localization:** Regulations should consider the unique risks and vulnerabilities of different regions and communities.
* **Behavioral:**
  + **Ethics:** Developers should prioritize safety and minimize potential harm in all stages of AI development and deployment.
  + **Morals:** Moral considerations should guide the development and deployment of AI systems, ensuring that they are used for beneficial purposes and do not cause harm to humans or the environment.
  + **Virtues:** Virtues such as compassion, empathy, and a commitment to the well-being of others are crucial for ethical AI development.
  + **Norms:** Industry norms should promote the development and deployment of safe and beneficial AI systems.
  + **Customs:** Cultural norms related to safety, risk avoidance, and the well-being of others should be considered when developing and deploying AI systems.
* **Conceptual:**
  + **Foundational Concepts (Rights, Responsibilities):** Rights (e.g., right to safety, right to be free from harm), Responsibilities (e.g., responsibility of developers to ensure the safety and well-being of others).
  + **Operational Concepts (Obligations, Privileges):** Obligations (e.g., obligation to minimize harm, obligation to conduct thorough safety assessments), Privileges (e.g., the privilege to use AI for beneficial purposes, subject to appropriate safety and risk management measures).
  + **Contextual and Evaluative Concepts (Consequences, Intentions, Duties, Expectations, Practices, Directives):** Consequences (e.g., the potential for AI systems to cause harm to humans or the environment), Intentions (e.g., the intended purpose of the AI system and its potential impact on human well-being), Duties (e.g., duty of care, duty to prevent harm), Expectations (e.g., public expectations regarding the safety and ethical use of AI), Practices (e.g., best practices for risk assessment and management, safety testing procedures), Directives (e.g., ethical guidelines for AI development and deployment, codes of conduct).
  + **AI Ethics:** Ensure that AI development and deployment align with ethical principles, minimizing unintended consequences and promoting societal benefit.
  + **Cultural Differences:** Consider the cultural context when evaluating potential harms and risks associated with AI systems.
  + **Personal Values:** Individual values related to compassion, empathy, and the avoidance of harm should guide personal interactions with AI systems.
  + **New Directions in Digital Ethics:** Explore the long-term societal and environmental impacts of AI, and develop strategies to mitigate potential harms.
  + **General Cognitive Abilities:** Reasoning & Problem Solving, Knowledge Acquisition, Metacognition, Adaptive Behavior, Abstract Thinking are crucial for AI to anticipate and mitigate potential harms.
  + **Genuine Reasoning:** AI should be able to engage in genuine reasoning about the potential harms of its actions and decisions, and strive to act in ways that minimize or avoid harm.
  + **Artificial Superintelligence:** In the context of ASI, the principle of non-maleficence takes on even greater significance. It is crucial to ensure that ASI systems are developed and deployed in a way that minimizes the risk of existential harm to humanity and the environment. This may involve developing robust safety mechanisms, ensuring alignment with human values, and carefully considering the long-term consequences of ASI development. Minsky's framework of multiple interacting agents suggests that ensuring non-maleficence may require careful coordination and control of individual agents within the system to prevent unintended consequences and avoid cascading failures.