

Overview of Risk and Ethics

Ken Baclawski

The Common Definitions of Risk

- International Standard
 - “Effect of uncertainty on objectives”
 - Includes both positive and negative effects
- Dictionary (OED)
 - “(Exposure to) the possibility of loss, injury, or other adverse or unwelcome circumstance; a chance or situation involving such a possibility”
 - Only includes negative effects

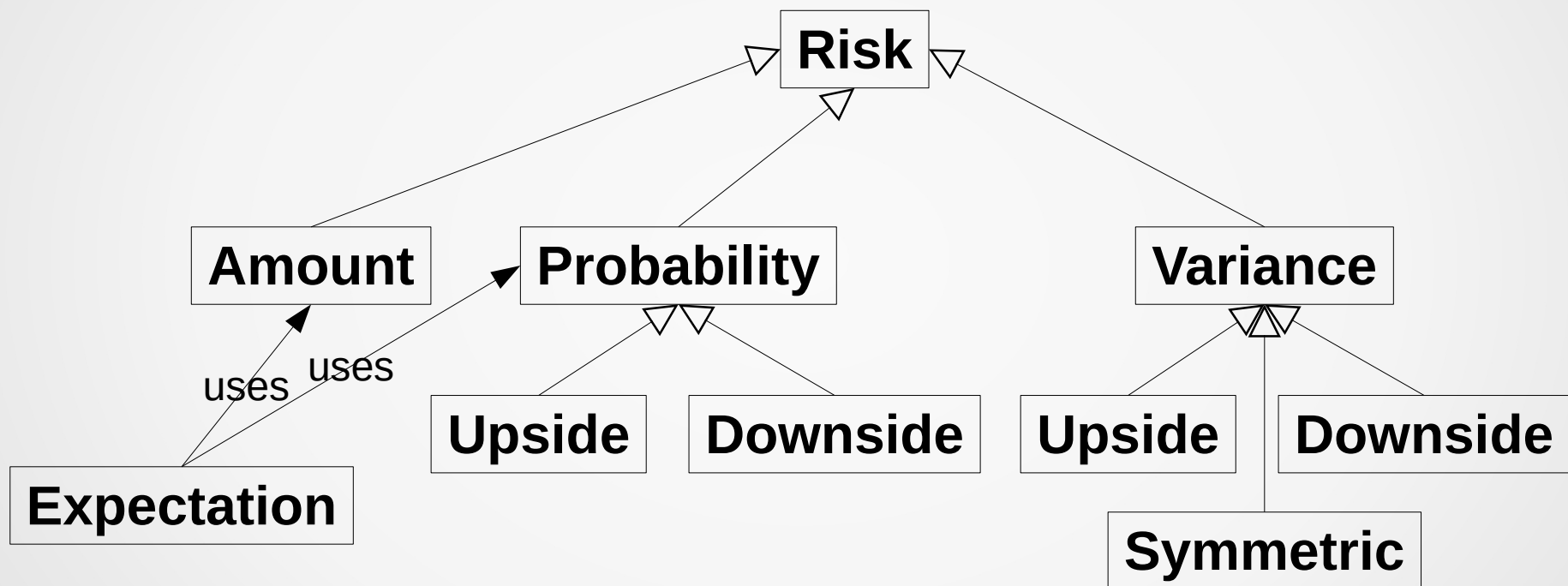
Simple example

- Consider the phrase “risk \$1000 on a venture”
 - What is the risk?
 - The amount \$1000 that is at risk
 - The probability that the venture will not achieve the expected return
 - Suppose that the amount was \$2000
 - First interpretation: \$2000 is twice as risky.
 - Second interpretation: The risk is the same.
 - The common definition is the second interpretation, but the first interpretation seems to be more “intuitive”.

Another Interpretation of Risk

- Consider the phrase “risk \$1000 on a venture” again
 - A “safe” venture is one with a low variance
 - A “high-risk” venture is one with a high variance
- Two ventures can represent different risks (using the interpretation of risk as a variance) even if they have the same amount at risk and have the same probability of achieving objectives.

Taxonomy of Risk



Potential AI System Risks

- Terminology
- Software engineering
 - Scale and complexity
 - Security
 - Privacy
 - Validation and testing
 - Application evolution
- Lack of explainability
- Lack of reproducibility
- Harmful bias
- Environmental impact
- Poor data quality
- Automation-spurred job loss
- Deep fakes
- Socioeconomic inequality
- Market volatility
- Weapons automatization
- Uncontrollable self-aware AI

UNESCO AI Ethical Principles

- Proportionality and Do No Harm
- Safety and Security
- Fairness and non-discrimination
- Sustainability
- Right to privacy and data protection
- Human oversight and determination
- Transparency and explainability
- Responsibility and accountability
- Awareness and literacy
- Multi-stakeholder and adaptive governance and collaboration

Potential AI System Risks Again

- Terminology
- Software engineering
 - Scale and complexity
 - **Security**
 - **Privacy**
 - Validation and testing
 - Application evolution
- **Lack of explainability**
- Lack of reproducibility
- **Harmful bias**
- **Environmental impact**
- Poor data quality
- **Automation-spurred job loss**
- **Deep fakes**
- **Socioeconomic inequality**
- **Market volatility**
- **Weapons automatization**
- **Uncontrollable self-aware AI**

References

- B. Hubert, Why Bloat is Still Software's Biggest Vulnerability, **IEEE Spectrum** April 2024, 22-25, 50
- K. Baclawski, M. Bennett, G. Berg-Cross, D. Fritzsche, R. Sharma, J. Singer, J. Sowa, R.D. Sriram, M. Underwood and D. Whitten. Ontology Summit 2019 Communiqué: Explanation. **Applied Ontology**. IOS Press, The Netherlands. DOI: 10.3233/AO-200226. (2020)
- K. Baclawski, What is Risk?, Ontology Summit 2022, <https://bit.ly/3ZAONIC>
- Recommendation on the Ethics of Artificial Intelligence, UNESCO, 23 November 2021, <https://bit.ly/4bhG5of>