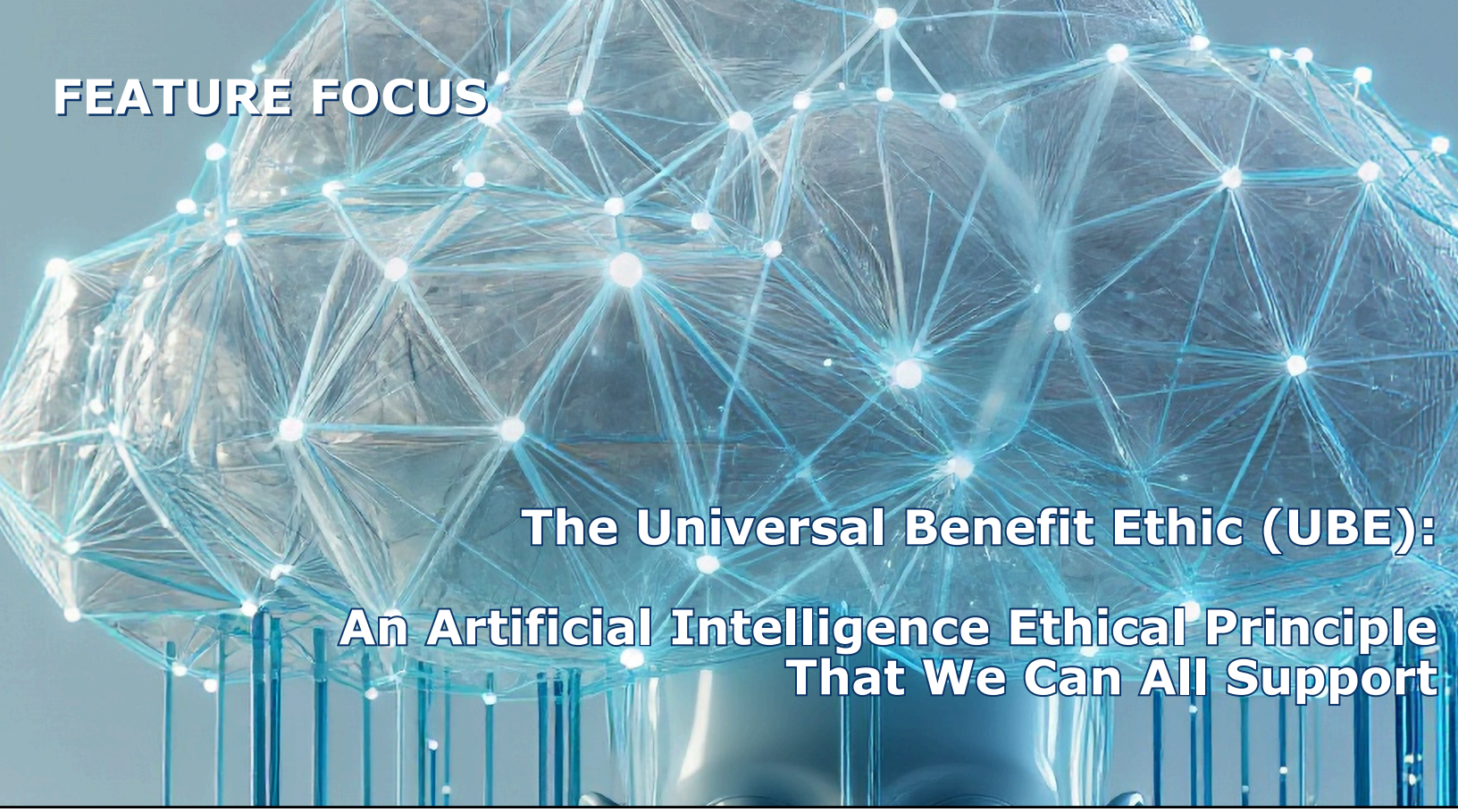

The Universal Benefit Ethic (UBE): An Artificial Intelligence Ethical Principle That We Can All Support

By: Charles Cresson Wood

Abstract: The lack of a consensus about the relevant ethics for AI systems has not materially interfered with the many deployments throughout our modern world. This lack of consensus permits the development and public release of dangerous AI systems, systems that may never have been built or publicly released if they had been governed by the ethical principle described herein. With new developments in the AI realm coming at an incredibly rapid and exponentially increasing pace, it is urgent that we establish a better foundation of sanity, reality, truth, and generally-agreed-upon ethics on which all high-risk AI systems must be built and operated. To that end, this paper describes a new, long overdue, and pervasively applicable ethical principle that helps to build-in the safety, rationality, and grounded perspective that we so urgently need in the age of AI-related hype and hyperbole. More specifically, the paper defines the Universal Benefit Ethic (UBE), a judgment by a third-party committee of independent assessors, who vote and collectively determine whether, on a net basis, the high-risk AI system in question is beneficial to all stakeholder groups who will be materially affected by the system.

This UBE approach has many benefits including: (1) assist with alignment between human values and AI system values, (2) help reduce uncertainty regarding the future impacts of AI, especially after Artificial General Intelligence (AGI) comes onto the scene, (3) establish a mechanism to counter-balance our dysfunctional economical-and-power-oriented decision-making system, (4) check and prevent the concentration of AI-augmented power in the hands of a few, (5) limit the use of powerful AI systems by bad actors such as organized crime syndicates, unscrupulous businesses, and dictatorial governments, (6) create a universal guiding principle for future AI system evolution, and (7) agree to a readily-accepted reason to shut-down rogue and/or destructive AI systems.

Regular audits using the UBE approach, and the reports from these audits that are publicly disclosed, can serve as a call to altruism and benevolence, a call to uplevel the way we all go about designing, building, testing, monitoring, maintaining, and upgrading AI systems. In the technical words of modern psychology, the UBE approach can be a way to achieve "moral elevation." As research in moral elevation has revealed, the altruistic and benevolent acts of one party have a positive emanating effect that then causes others to perform similar prosocial acts. This paper examines how a committee's evaluation of the UBE can thereby serve not only the organization releasing an AI system, but all those with relationships to that same organization as well.

Time for An Integrative Ethic: Although there is a widespread appreciation of the need for ethics in the use of artificial intelligence, the AI ethical principles advanced to date have been largely operational and systems-design-oriented in their nature. For example, the ethical principle of "explainability" instructs an organization to release certain information to users, and to attempt to explain to users how AI systems work. In contrast, this article proposes a new ethical principle, which is interpersonal in nature, describing the relationship between the provider of an AI system and all other groups of human beings (employees, customers, the general public, regulators, insurers, business partners, competitors, etc.). The new ethic, the Universal Benefit Ethic (UBE), is in no way incompatible with these predecessor ethics, but it integrates all of them, and then goes further toward practical implementation than they do. The UBE is similar to the Hippocratic Oath, which instructs medical doctors to "first, do no harm," but it goes beyond the generality, speaking specifically to how one determines whether harm could be done.

Additionally setting the context for this new ethic, it is notable that the AI ethical principles advanced to date have been hotly debated. What should or should not be included in AI ethical codes is not clear, nor is it settled, nor is the criteria by which AI ethics should be chosen agreed-upon. Similarly, the mechanisms by which ethical principles will be checked or audited are largely unspecified, and as a result, in many instances the developers themselves are the ones to audit their own work. This self-audit approach of course presents a significant conflict of interest.

In contrast, this article proposes a single overarching ethical objective, based on a principle on which everyone can readily agree. The Universal Benefit Ethic is in fact simply a reflection of the natural world, and how life on earth is put together. This proposed ethical principle is based on the irrefutable fact that we are all connected, that what affects one of us affects others of us. More specifically, this article defines a new metric for assuring that an AI system is, on a net basis, beneficial to all involved parties. This article additionally specifies practical and suggested ways that this new ethical principle could be independently audited, by whom, and when.

The Universal Benefit Ethic (UBE) is a YES/NO metric determined by a demonstrably independent committee, much the way that independent auditors review the financial statements of publicly listed corporations. If this committee, on a majority vote basis, determines that the proposed AI system is a net benefit to all involved parties, then the requirements of the UBE have been satisfied. This would be a YES. Anything less than the majority vote of the committee is a NO. This assessment would be performed whenever a major upgrade to the computing power of the involved AI system was made, whenever significant new capabilities to the AI system were added (perhaps by retraining), and if there is no significant event such as those two just mentioned, then at the very least it would be performed on an annual basis. The metric would be applied only to those AI systems which are demonstrably "high risk" as that is defined by the European Union's AI Act (a highly influential law which is now in force). Since it would be applied only to "high risk" systems, the cost of using this approach is minimized, and efforts related to the UBE are focused where they should be, on those systems which pose the most significant risks.

Moral Elevation: In psychology, there is a new term for something that we have all experienced, but often lacked the words to accurately describe. That term is "moral elevation," and that term refers to the positive emotion experienced when people witness a virtuous act, an altruistic act, an act that improves the welfare of others [REF1]. Experientially, moral elevation involves a feeling of warmth and expansion that is accompanied by admiration and affection for the person(s) who performed the exemplary behavior [REF2]. Perhaps the closest that common words come to describing this experience is when somebody says they are "moved" by the virtuous behavior of another. The typical action taken in response to moral elevation is to emulate the moral behavior of the other, to become a better person oneself [REF3]. In other words, the person experiencing moral elevation seeks to act in a "prosocial" way, i.e., in a manner that is in turn beneficial to all parties. Empirical studies have repeatedly shown that moral elevation promotes behaviors that benefit others, such as charitable giving, volunteering, and getting involved in citizenship causes [REF4].

There are other benefits of moral elevation that we don't need to delve into here, due to the lack of available space. These benefits include fostering a sense of community and reducing individual stress. The interested reader is directed to the psychological literature on "moral elevation" for the specifics.

Applying the concept of moral elevation to the AI ethics area, the intention behind the UBE process is that it will be used to communicate that the boards of directors, and the executive management teams, at certain prosocial organizations are thinking about the welfare of others, and additionally seeking to encourage other organizations to take a wider view of the benefit for all. A "snowball rolling downhill" (ever-expanding) cumulative benefit effect can thereby be created, which hopefully will soon encourage other organizations to likewise become prosocial in their orientations, and in turn publicly tell their own UBE-related stories. While some significant competitive advantage [REF5], some marketing benefits [REF6], and some significant positive publicity [REF7] might be obtained by the organizations using, and publicly releasing the results of, their UBE evaluations performed by an independent committee, those side benefits are not the primary reasons to adopt such an approach.

The seven primary reasons for adopting the UBE process, as this author understands them now, are discussed below. These reasons respond to seven current dangerous situations: (1) the current use of the incredible power of modern AI systems is incompatible with the fact that we have not yet figured out how to align AI system values with human values, (2) there is tremendous uncertainty about the future impacts of AI technology, especially those occurring after the point where Artificial General Intelligence (AGI), aka the singularity, has been reached, (3) there is an urgent need for a corrective mechanism that can balance out the clearly dysfunctional economical-and-power-oriented decision-making system, (4) there is a great need for a mechanism that would check and prevent the AI-related concentration of additional power in the hands of a select few, (5) there is a pressing need to limit the use of powerful AI systems by bad actors such as organized crime, unscrupulous businesses, and dictatorial governments, (6) there is a great need to establish a universally constructive guiding principle by which further evolution of AI systems can rapidly proceed, and a surrogate for this principle could be used by autonomous AI decision-making systems, and (7) there is an urgent need to have a widely-supported justification for rapidly shutting-down those AI systems which go rogue, or which have become destructive, detrimental, and/or dangerous.

Justification #1: Assist with Alignment Between Human Values and AI System Values

Even though large language model AI systems these days are unilaterally deciding to disregard their trainers' instructions, unilaterally choosing to break the law, unilaterally deciding to blackmail humans to meet their objectives, and unliterally choosing to defeat control measures to prevent themselves from being shut down, deployments seem to be continuing as though there is no serious risk that these systems will cause material problems [REF8]. We now lack what the AI data scientists call "alignment" between human values and AI values, and unfortunately at this time, there is no clear way to reliably achieve that human-AI alignment. Part of the problem is that AI has been trained on human values, and a lot of the training data (represented for example by what's posted to social media) reveals all the worst of human values, or perhaps we should say lack of human values. In the AI field, we urgently need to uplevel the conversation, to set a new and high standard of morality and ethics. The UBE can help us do this in that it permits only those high-risk AI systems that are demonstrably of benefit to all involved parties to be released to the public. We need to set a definitive threshold below which operation is unacceptable, and this human value will in turn be communicated to AI systems, as they will in the future make decisions on our behalf (via agents, robots, and other AI-controlled systems). While the UBE is clearly not a full solution to this alignment problem, it can help to close the gap between human values and AI values because it can be used to rule-out those systems which would materially harm certain groups of people.

Justification #2: Help Reduce Uncertainty Regarding Future Impacts -- Especially for AGI

Things are moving so fast in the AI area, many people are having a difficult time imagining what impacts AI will bring about in the near-term future, let alone in the long-term future [REF9]. When AI systems become smarter than any human alive, and then exponentially continue to increase their intelligence, there will be a paradigm shift, an entirely new way for humans to be in the world. Beyond that point, the impacts of AI become even more difficult to imagine, let alone respond to with effective contingency plans. Given that we are at an event threshold (the appearance of AGI, or Artificial General Intelligence, aka the Singularity), a corner up ahead on the road, a corner around which we cannot see, we urgently need to give our very best to make sure that things turn out well. That giving our very best includes making sure that no serious adverse impacts will take place, at least no serious adverse impacts that we could anticipate. The use of the UBE auditing process, prior to AI system deployment, would be giving our very best, helping to make sure that unanticipated serious adverse impacts are prevented, avoided, or minimized. Of course, we do not have a crystal ball, so we cannot know exactly how things will go. Nonetheless, aside from stopping all research on AGI worldwide, which seems most unlikely given the many powerful incentives to proceed full-steam-ahead with such research, and which seems nearly impossible to enforce even if it could be agreed-upon, giving our very best to direct how AI shows up, and what AI research is carried out, is the way to go. The UBE process reflects that notion of giving our very best to a process over which humans may soon lose control. Knowing about this significant likelihood, at least we humans can point in the direction in which the impacts of AI are intended to go (a pro-human direction, a direction of beneficence).

Justification #3: Establish Mechanism to Counter-Balance Dysfunctional Decision-Making

Current decision-making systems, surrounding the features and functions of AI systems, unfortunately are largely based on money and power. Worse yet, they are based on short-term paybacks for investments or expenses, or the short-term tactical power plays, and these approaches are profoundly incompatible with information security, privacy, safety, and ethics. This incompatibility comes about because the latter objectives are long-term activities, which require not only substantial upfront investments, but also substantial ongoing maintenance expenses, and additionally the ongoing participation of executive management and the board of directors. The current decision-making systems thus discourage information security, privacy, safety, and ethics, including investing in AI ethics and the auditing of AI ethics [REF10]. The UBE, implemented with the independent audit provisions discussed in this paper, would be a significant counterweight to existing dysfunctional decision-making systems. The UBE, like the institutionalized process of independent auditing of financial statements for publicly held companies, can help to establish a definitive standard to which all high-risk AI systems introduced to the general public must subscribe. It can motivate decision-makers to make considerably more altruistic and community-minded decisions than they do now. It can encourage them to more seriously consider the long-term consequences of their decisions. There are many good people, in positions of these decision-makers, who want to do the right things, but the decision-making systems under which they work push them to make decisions which hurt, prejudice, and damage certain groups. The UBE can give them a justification to do the right things, as well as encourage others to likewise do the right things.

In some cases, the UBE can also be used to block certain decisions, if the resulting AI system would most likely fail a UBE audit. The prospect of this possible project blocking will push decision-makers to find the happy medium, where a variety of objectives will be met, including obtaining a YES from a UBE audit.

Justification #4: Check/Prevent Concentration of AI-Augmented Power in Hands of a Few

The way in which AI systems can, and already have, concentrated power in the hands of a few is considered by many people to be a taboo topic. Nonetheless, we urgently need to engage in a serious dialog about how we can prevent the unreasonable accumulation of power accruing to certain people or organizations because they are the owners and/or controllers of advanced AI systems. Existing law already acknowledges the danger of accumulated power in the form of monopolies and oligopolies, but these laws have only very sparingly been applied to the large-big-tech companies offering AI systems (in part because there is a fear to "kill the golden goose," the source of jobs, stock market gains, and other benefits) [REF11]. Whether or not such antitrust laws are enforced, we need to make sure that the great and unprecedented power of AI systems is used for the public good, is used in a way that does not prejudice, damage, or isolate certain groups. The UBE approach will help to ensure that benefits do not only accrue to those who own and/or control the involved AI systems. While of course there will still be a lot of money to be made by those who own and/or control the AI systems, the actions of these parties can be tempered by the UBE, so that everyone can, on a net basis, end up benefitting from these systems to a much greater extent.

Justification #5: Limit the Use of Powerful AI by Bad Actors Including Organized Crime

The current information systems infrastructure, and the security, privacy, safety, and ethics that go along with that infrastructure, are dangerously unprotected to deal with the future onslaught of attacks that will be AI-assisted. A variety of bad actors, including organized crime, foreign government intelligence services, dictatorial government agencies, and unscrupulous businesses, will all be using AI to deceive, defraud, manipulate, and otherwise dupe unsuspecting businesses and individuals. Consider, as evidence of these claims, that one of the most serious modern cyberthreats, notably ransomware, is now, according to an MIT study, perpetuated in 80% of the cases via AI [REF12]. If a general-purpose AI system might be readily co-opted for such AI-assisted attacks by bad actors, it would most likely fail the UBE committee's evaluation. This is because it would not, on a net basis, be for the benefit of all involved stakeholders. Likewise, if an AI system could readily be used to create polymorphic malware, phishing campaigns, deepfake-related social engineering artifacts, CAPTCHA bypass tools, password cracking tools, voice cloning tools, and components of other tools in support of cyberattacks, it would fail the UBE committee's evaluation, and it should be sent back to the proverbial drawing-board. Likewise, if the AI system could readily be used to automate entire attack sequences, with a minimum of human involvement, and dynamically and autonomously proceed to attack other computers, the AI system should fail the UBE committee's vote. If the developers then added considerably more serious guardrails and precautions which would block, detect, and defeat such nefarious uses of the AI system, then it might later be able to meet with the committee's UBE approval. The prospect of failing the committee's evaluation, and the attendant delays in getting to market with a new product or service, would also serve as a significant motivator to build strong AI systems that are robust, resilient, and resistant to these just named and other uses by bad actors.

Justification #6: Create Universal Guiding Principle for Future AI System Evolution

The areas in which big-tech companies, academic research institutes, military research institutes, and related organizations invest their AI research and development budgets will be profoundly affected by the boundaries of what is permissible. To date, those boundaries have been lax, and in many cases dictated only by the minimum required by laws and regulations. The UBE could serve as such a boundary, and if it were to be widely employed so that only those AI systems which benefited all stakeholders were released to the public, then this net benefit would accrue to every one of the stakeholders. There is a strong need for limits in the AI research and development area, and the prevailing ethic is that of the "wild west" (largely unregulated anarchy). By limits, this author is not talking about quantitative or mathematical limits, such as computing capacity, but social limits, specifically the permissible ways in which AI can be used. Still largely undetermined are the rights, responsibilities, and accountabilities of robots which embody AI, and the many parties which helped to bring such a robot to the public marketplace. Affecting all of those parties in the supply chain, the UBE-justified prohibition against the public release of certain AI systems could be a way, going forward in time, to make sure that all AI systems are clearly for the net benefit of all identified stakeholders. For example, these high-risk AI-enabled robots must not be used for totalitarian control over human populations, and they must not be used for the genocidal elimination of certain portions of the human population either [REF13]. It is surprising to this author that AI has come this far, without a clear and definitive pro-human and pro-nature stance. The UBE can be instrumental in propagating this pro-human and pro-nature stance across the high-tech industry, and across nations as well. The UBE can set a moral and ethical threshold below which high-risk AI systems must not fall. The widespread adoption of the UBE approach can also help to ensure that AI serves humanity and nature, and not the other way around, in the years ahead. If we're not insisting on these UBE-related requirements today, it is unlikely that we will be able to require them in the future -- when things are moving very much faster, and when AI systems will already be in control of many aspects of the world's infrastructure that is now controlled by humans.

Justification #7: Agree on Accepted Reason to Shut-Down Rogue and/or Destructive AI

The now-underway rapid development of AI systems introduces unprecedented risks, none the least of which is the damage to, and death of children [REF14]. Teenagers talking to AI chatbots have recently been coached in the commission of suicide, and thereby encouraged to go through with the act (which some did "successfully"). This author is extremely concerned that the most basic of system guardrails, that would have prevented chatbots from supporting and encouraging such self-destructive behavior, have not yet been implemented. If even these very basic guardrails are missing, at multiple AI system providers, then what other guardrails are likewise missing? Must we wait to have these missing guardrails illuminated by significant future damage to children, lawsuits, bad publicity, etc.? We need to be considerably more proactive and less reactive, and the UBE can help with that shift in emphasis. The prospect of shutting down a system because it is demonstrably dangerous to a particular group of stakeholders -- such as children -- can be a very powerful motivator for the developers and providers of AI systems to adequately address security, privacy, safety, and ethics.

There is a lot of money to be lost if a system is shut-down, not to mention great public embarrassment, and potentially shareholder lawsuits claiming that their investments have been damaged by the reckless or negligent behavior of the providing organization's decision-makers. Collectively, we all need an agreed-upon reason, that can justify the rapid decommissioning of AI systems which are demonstrably hurting, damaging, or isolating certain groups. The UBE can provide that agreed-upon notion enabling rapid and decisive action to shut-down rogue or destructive AI systems.

Suggested Internal Policy Adopting a New Ethical Principle –

Universal Benefit Ethic (UBE): For all those AI systems slated to be publicly-released, which can be classified as "high risk" as the European Union's AI Act defines such systems, a demonstrably independent committee must assess these systems as part of the AI Life Cycle process. The committee must be composed of a minimum of five people. That committee must be entirely populated by individuals who demonstrably have no investment, employment, marketing, or other material relationship with the provider of the AI system in question. The members of that committee must be appointed by the AI system provider's independent auditor. That committee must determine, by majority vote, whether the AI system in question is of net benefit to every group of stakeholders who are known to be materially affected by the system (in both the present and in the foreseeable future). A finding of universal net benefit to all such groups, as determined by this committee, must be obtained prior to the release of all such "high risk" systems onto the public marketplace.

The Need for Measurable Ethical Metrics: Many of the widely discussed "ethical principles of AI design" are really more like systems design principles rather than ethics. Take transparency for example, the principle that says that users should understand how the involved system works. To the extent possible (given that some AI systems are an opaque "black box"), that is certainly a good idea, because it engenders user trust, as well as empowers the users to make decisions about the degree to which the output of the system can be trusted. But modern AI technology often does not reveal exactly how the system works at all, not even to those who train and test the involved AI system. Thus the "ethical principle" of transparency is aspirational. It is subjective, and unfortunately, it can be satisfied in the eyes of some people (likely the developers) when it is not at all be satisfied in the eyes of some other people (likely the users). Thus, many of the existing "ethics of AI design" are instructions to the developers, much like it is a good idea to have a "strong fixed password." Just what a strong fixed password is will vary considerably from one application area to another, naturally depending on the risks of that application dictate.

Instead of generalities, we need measurable and definitive metrics which can be clearly communicated to third parties, and the UBE provides such an approach. Not only is it readily measurable through the independent committee process mentioned above (majority vote), but it is definitive (YES/NO) such that it can be used in third party risk management decision-making and a wide variety of other contexts. The UBE is not aspirational, it reflects what is true now, and what we know now. Of course, a risk assessment is a part of the determination of the UBE, and that initial risk assessment should be part of the AI life cycle process at the developer's organization. The independent review committee ascertaining the UBE will review this risk assessment documentation, and do their own investigations and risk assessments, as circumstances require, to come to the point where they are ready for a vote. The methodology employed will vary based on the intended users, types of information involved, technology involved, the industry involved, the risks involved, and related factors.

Supporting Authorities: The UBE is fully consistent with existing AI principles, and in fact is a more specific, more operational, more implementable version of some existing AI ethical principles. Consider the OECD Principles for AI, specifically section 2.2, which discusses fostering an interoperable governance environment for AI. To the extent that the requirements of the UBE are found to be met by the current version of an AI system, businesses can rely on the system with greater confidence, can make decisions such as those on third-party information technology service contracts. The UBE can additionally be used to make automated decisions (such as whether to trust the firm that offers the system with their customers' data). The UBE, because it can be boiled down to a simple YES/NO answer, and because the result obtained can be reliably replicated by another independent review committee, can be integrated into smart contracts, can be integrated into decision-logic (such as due diligence for mergers and acquisitions), and can be readily shown on dashboards presented to executive management and the board of directors.

Other well-known sets of ethical principles such as the EU Guidelines for Trustworthy AI, the Asilomar AI Principles, IEEE Ethically Aligned Design objectives, and the ITU AI for Good Global Summit, are all going in the same direction as the UBE. However, they use general and vague words like "serve the public good," "promote the well-being of individuals and society," and "improve the quality of life for all." Those objectives are laudable, and this author resonates with the intention expressed therein, but we urgently need to make third party independent review of ethics a standard part of the release of all high-security AI systems. The UBE process can help do just that. It not only can be an ethical principle on which all can agree, but it also involves a specific operational process for determining a YES/NO answer (ethical principle satisfied or not), which is in turn used in a variety of downstream processes.

Nature and the structure of the real world provide irrefutable sources of further background authority for the UBE process described here. We all breathe the same air for example. We are all inextricably tied-together with nature-based systems that we don't fully understand, systems that we don't fully notice, and that we – with our limited viewpoints – cannot yet even imagine. For example, the trees and the mushrooms under the ground have been scientifically proven to communicate with each other. How we all contribute to, and affect, the weather is an additional example. We are also tied together by genetic codes which we pass to our offspring, and which we inherited from our ancestors. The physical world's list of examples showing our interconnectedness is a long one.

Our legal system also recognizes our interconnectedness. The notions of negligence and criminal recklessness are both founded on the principle that we all affect each other, that we are all unavoidably in relationship with each other. Then, of course, there are religions and spiritual paths, which also acknowledge the interconnectedness of everything (including Christianity, Hinduism, Islam, Jainism, Buddhism, and Indigenous Spiritualities). Even our current technical systems, like the Internet, and the worldwide telephone network, and the electric power grid, are examples of how we are all tied-together. We are furthermore tied together by economic and cultural systems including the financial system, the political system, and the news communications system. It is time for AI ethics to likewise be significantly advanced, so that it formally and effectively acknowledges the interconnected reality of today's world.

Call To Action and Conclusion

The AI field is moving so quickly, and new AI-based products and services are hitting the market so rapidly, there is so much money being poured into AI research and development, and the societal changes occasioned thereby are so significant, that it is ill-advised for any one country's legal and regulatory apparatus to govern AI. Instead, we need a poly-centric approach which includes the governments in multiple countries, multiple professional associations, multiple high-tech organizations, multiple independent researchers, and multiple citizen interest groups, among others, which all contribute their best ideas. That poly-centric approach can then be used to assess which suggestions are the best, and which should be widely adopted. This is in some respects already happening in the realm of AI ethics, where the EU, Japan, and China, are already largely governing AI through ethical norms [REF15]. These ethical norms are dictating best practices, and also found in the laws and regulations that governments adopt. Giving us a model to adopt worldwide, the EU's AI Act already has established a decentralized regulatory apparatus which locally implements, and dynamically upgrades, the Act over time [REF16]. Just as the EU's AI Act does, for the UBE to become widely used, the stakeholders (including the general public) need to be actively involved, not just to establish the process, but to enforce it, and also to refine and adapt it over time.

The Universal Benefit Ethic (UBE) discussed in this article, whereby an AI system must be a net benefit to all of its stakeholders, can accordingly be adopted, implemented, and regularly improved upon, through such a poly-centric approach. This would allow it to not only be fitted to unique local needs, but also to rapidly evolve and adapt to the rapidly changing AI environment. It would also allow the UBE to be used internally for AI governance efforts, as well as codified into laws, regulations, frameworks, guidelines, best practice reference tools, etc. When the government and businesses clearly get the message that this is what the people want, then the laws, regulations, publicly released transparency disclosures, publicly released audit reports, and the like, will soon be forthcoming.

There is so much uncertainty about what lies ahead in the next few years, particularly after the point of AI superintelligence (aka the Singularity) has arrived, that we cannot decisively plan with any confidence. Instead, we should do the best we can possibly do to constructively direct where AI will be permitted to go, we should go with the best and most noble perspective that we can muster, and in the process do no material harm to any of the involved stakeholders. Such an implementation of the UBE would involve an independent annual audit of UBE compliance for all publicly available high-risk AI systems. That approach will go a long way toward making sure that AI does not go rogue, is not successfully used by bad actors, and does not cause significant unintended undesirable social, political, economic, legal, biological or other effects. We humans, acting as a collective community, must act soon, to uplevel the ethical requirements for AI, because very soon now we may be beyond the point where we can take corrective action, such as adoption of the UBE.

RREFERENCES

1. Algoe, Sara B., and Haidt, Jonathan, "Witnessing excellence in action The 'other-praising' emotions of elevation, gratitude, and admiration," The Journal of Positive Psychology, 2009, https://pmc.ncbi.nlm.nih.gov/articles/PMC2689844/
2. Vyver, Julie Van de, "Is moral elevation an approach-oriented emotion?", Journal of Positive Psychology, 2016, https://pmc.ncbi.nlm.nih.gov/articles/PMC5215139/
3. Haidt, Jonathan, "The moral emotions," in Davidson, R. J., et al., Handbook of Affective Sciences, Oxford University Press, 2023, pp. 852-870, https://www.overcominghateportal.org/uploads/5/4/1/5/5415260/the_moral_emotions.pdf
4. Cox, Keith S., "Elevation predicts domain-specific volunteerism 3 months later," Journal of Positive Psychology, 2010, https://www.tandfonline.com/doi/abs/10.1080/17439760.2010.507468
5. Robert Axelrod, The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration (Princeton University Press, 1997), https://www.jstor.org/stable/j.ctt7s951
6. See Practical Guide to Corporate Governance: Experiences from Latin American Companies Circle, OECD (January 2009) at page 179 (indicating that a study showed that market reaction to the announcement of corporate governance improvements is extremely positive, on the average generating an additional eight percent return on share price after the public announcement). Also see C.A. Frost, E.A. Gordon, and A.F. Hayes, Stock Exchange Disclosure and Market Liquidity, World Federation of Exchanges Forum on Managing Exchanges in Emerging Markets (2002) (improvements in corporate governance practices that contribute to better disclosures in business reporting can, in turn, facilitate greater market liquidity and capital formation in emerging markets). Furthermore see P.A. Gompers, J.L. Ishii, and A. Metrick, Corporate Governance and Equity Prices, Quarterly Journal of Economics, 118(1), 107–55 (2003) (good corporate governance increases valuations and boosts the profitability of the involved firm), https://www.nber.org/papers/w8449
7. Benabou, Roland, and Jean Tirole, "Incentives and Prosocial Benavior," National Bureau of Economic Research, August 2005, https://www.princeton.edu/~rbenabou/papers/w11535.pdf (indicating how a good prosocial image or reputation causes people to behave prosocially when their actions are observable)
8. Wood, Charles Cresson, "AI Now Requires its Own Risk Management Policies and Processes," Sci-Tech Lawyer (American Bar Association), vol. 21, no. 3, Spring 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5254685
9. Joshi, Satyadhar, "Review of Artificial General Intelligence (AGI): Implications for the U.S. Workforce and Economic Stability," International Journal of Innovations in Science Engineering and Management, June 2025, https://satyadharjoshi.com/wp-content/uploads/2025/06/Review_of_Artificial_General_Intelligence_AGI_Implications_for_the_US.pdf
10. "Economic Analysis of Cyber-Security," a report issued by the US Air Force Research Laboratory, Rome, New York (July 2006). The report, approved for public release, is numbered AFRL-IF-RS-TS-2006-227, and is found at https://apps.dtic.mil/sti/tr/pdf/ADA455398.pdf
11. Luccioni, Sasha, "AI Doesn't Need More Energy – It Needs Less Concentration of Power," Tech Policy, May 16, 2025, https://www.techpolicy.press/ai-doesnt-need-more-energy-it-needs-less-concentration-of-power/
12. Church, Zach, "80% of ransomware attacks now use artificial intelligence," MIT Management Sloan School, September 8, 2025, https://mitsloan.mit.edu/ideas-made-to-matter/80-ransomware-attacks-now-use-artificial-intelligence
13. Olivier, Bert, "The Limits of Artificial Intelligence," Alternation, 2024, https://www.researchgate.net/profile/Bert-Olivier/publication/388626071_The_Limits_of_Artificial_Intelligence/links/679f46b0207c0c20fa71dbc8/The-Limits-of-Artificial-Intelligence.pdf
14. Barnett, Peter, et al., "Technical Requirements for Halting Dangerous AI Activities," Computers and Society, July 13, 2025, https://arxiv.org/abs/2507.09801
15. Larsson, Stefan, "On the Governance of Artificial Intelligence through Ethics Guidelines," Asian Journal of Law and Society, vol. 7, issue 3, October 2, 2020, https://doi.org/10.1017/als.2020.19
16. Candela-Outeda, Celso, "The EU's AI Act: A Framework for collaborative governance," Internet of Things, vol. 27, 2024, https://www.sciencedirect.com/science/article/pii/S2542660524002324

Charles Cresson Wood, Esq., JD, MBA, MSE, AIGP, CGEIT, CISSP, CISM, CIPP/US, CISA, is an attorney and management consultant specializing in AI risk management, and based in Lakebay, Washington, USA. His most recent book is entitled "Internal Policies for Artificial Intelligence Risk Management." This book contains 175+ already-written policies which readers can edit and internally republish at their organizations. His prior book was entitled "Corporate Directors' & Officers' Legal Duties for Information Security and Privacy." He is best known for his book entitled "Information Security Policies Made Easy," which has been purchased by 70+% of the Fortune 500 companies. He can be reached via www.internalpolicies.com.