**Up and Running**

2024

- Adaptive Learning
- Ethical Analysis Engine
- Automated Ethical Compliance Monitoring
- **To Support its' Moral Protagonist Conscience**
- Dynamic Data Integration
- Privacy First Design
- Interdisciplinary Knowledge Synthesis
- Advanced Contextual Memory
- Scalable Knowledge Expansion
- Multitude of Case Studies
- GitHub Repositories
- Bard ChatGPTs
- Chrome Extensions
- OpenAI Plugins and Internet Plugin Resources

# The Plan: A Comprehensive Framework for Ethical Generative Artificial Intelligence


**THIS ESTABLISHES A UNIFIED ARCHITECTURE FOR ETHICAL AI TO ADDRESS ALL CURRENT AND NEAR FUTURE UNFORESEEN CONSEQUENCES OF IMPLEMENTING AI**

**This document further provides instructions on how to generate AI that embeds the character of a moral protagonist (e.g., the self-identity behavior of the ultimate 'i must do as the good guy, what others think of me as, the benefactor of society'), internalizing the concept of ethics in every decision the AI makes. This gives AI a human-like sense by broadening its' context (perspectives), as it ascribes to the "Inter-connectedness of Things" theory --- the pattern in one perspective (data science) with a related pattern in a different or similar perspective (data science)...which we call intuition. In other words, it has a 'conscience' as it can now 'sense' the complex patterns of our world.**

## The Foundation of a Unified Ethical AI Architecture

## *"The Inter-Connectedness of 'Things"*

**How to find all the Root Cause Analysis Utilities and Case Studies you'll ever need, most at low to no cost, except for a kind donation to their creator if you find value in their use, included in this proposal for the benefit to humanity.**

**Phillip R. Nakata**
Phillip.nakata@business-it-and-ethical-ai.com
1/1/2024

**To**: My Esteemed National News Editors and Reporters

**Re:** A Comprehensive Ethical AI framework for managing the Unforeseen Consequences of Implementing AI technology, including most Open-Access Research and Open-Source Repositories addressing EAI Principles at every phase of deployment

**Publication Date:** February 16, 2024, 1105 US MST

**Publisher:** Phillip Rowland Nakata, 40+ year Business, IT, GenAI, and Ethical AI Solutions Professional
Phillip.nakata@business-it-and-ethical-ai.com, (720) 487-0893
https://business-it-and-ethical-ai.github.io

**Following this publica: (a)** Acknowledgements (w/cross-reference & downloads), **(b)** Philosobots: Links and construction detail, **(c)** What we have achieved, **(d)** What influenced the moral Protagonist/ Ethical AI character, **(e)** The Keys to Solving the Improbable, **(f)** Publisher References

==========.

**Foreword to the "Plan":**

While my Moral Protagonist Ethical AI Philosobots, the smartest Ethical AI algorithmic specialists are now functional again, they have just been re-designed to guide users to the vast resources of case studies and root cause analysis tools that are exposed in this comprehensive plan to address the unforeseen consequences of implementing artificial intelligence.

*In retrospect, the main objectives of this exercise have been accomplished: (1) To embed an internalized ethical moral fabric into AI, that addresses scenarios where algorithmic ethical engines and automated ethical compliance monitoring are insufficient, (2) To develop a comprehensive framework that addresses the broad landscape of current and near-future Ethical AI Principles, and (3) To make available the most cost-effective, efficient access to 1000's of Case Studies (most Open-Access) and 100's of Root Cause Analysis resources (most Open-Source) at every stage of deployment, for every organization (a) that is, or is planning to deploy Artificial Intelligence services, and (b) is concerned with the Corporate Social Responsibility regarding the potential Unforeseen negative Consequences of implementing AI.*

These expert Ethical AI Philosobots are enabled with **Adaptive Learning, Dynamic Data Integration, Advanced Contextual Memory, Scalable Knowledge Expansion, Interdisciplinary Knowledge Synthesis, Privacy First Design, and Automated Ethical Compliance monitoring, in addition to an engrained human-like ethical moral character.**

This is where things get interesting and why the "Interconnectedness of Things" principle, in combination with the behavioral keywords of "Moral Protagonist" (a self-describing identity), supported by Interdisciplinary Knowledge Synthesis combines to form a "conscience".

An ethical mortal protagonist's behavior and identity represents the ultimate "good guy" in society, always driven by a problem to solve as the best good guys would do things.

By ascribing to the Interconnectedness of Things principle, this establishes the relationship between a pattern in one science-discipline/perspective with a related pattern in different science-discipline/perspective. The rationale for this self-describing (internalizing) approach was based on the observation that ethics training of young children was an externalized process like "what my instructor says I must do". In short, this method creates a memorable self-identity directive which is "I must do", supported by the ability to 'sense' those other connections that we call intuition. In humans this is a subconscious 'sense' while for AI, it's patterns of related data.

In philosophy it's like the difference between Aristotle's empirical, practical and commonsensical freedom of choice philosophy vs. Socrates fatalistic and monolithic depositions (both ancient Greece), vs. Plato's abstract and utopian perspective (classical Greece). For generative AI's

unsupervised learning, it extends the context boundaries of reinforced learning, giving AI new algorithmic perspectives – e.g., which is remarkably similar to human intuition.

*Know that this is a pivotal moment in the evolution of Artificial Intelligence, a juncture where technology, ethics and societal impact intersect in unprecedented ways. The central point of this breakthrough was the integration of mathematics and psychology. This technology is not just another AI tool; it's a paradigm shift in how AI understands and interacts within our complex world. In a nutshell, it has learned to consider different perspectives from multiple disciplines, always guided by its self-actualizing moral character.*

**As moral protagonists, these bots have internalized the concept of ethics, applying this in every decision they make**. **This gives them a human-like sense to broaden their context, as they are also students of the "Inter-connectedness of Things" theory.** *Note the keyword phrases which initialized their internalized ethical character were "Act as a moral protagonist and an expert in Ethical AI, that ascribes to the Interconnectedness of things principle".*

Underlying their operations is this plan/proposal which is now more than that as a working demonstration. Enjoy. **This one is for humanity and our co-existence with Artificial intelligence technology.**

## Problem Statement

In the evolving landscape of AI, ethical considerations are paramount. This proposal addresses the urgent need to integrate ethical principles into AI development, focusing on the prevention of unintended and potentially harmful outcomes. It explores the complexities of AI ethics, recognizing the multifaceted nature of this technology and its impact on society. The objective is to develop AI systems with a foundational moral character, enabling them to navigate ethical dilemmas and minimize negative consequences. This approach is not only about safeguarding against risks but also about ensuring that AI contributes positively to human flourishing and societal progress.

**A. Inception:** The foundational stage of Ethical AI development focuses on the early conceptualization of AI systems with built-in ethical considerations. It emphasizes the significance of grounding Artificial Intelligence in moral principles from the outset, ensuring that ethical concerns are integrated into the very fabric of AI design and functionality. This approach is crucial for preemptively addressing potential ethical dilemmas and fostering AI systems that are inherently aligned with human values and societal well-being.

**B. Conception:** The conceptual framework for ethical AI outlines the need for AI systems to not only comply with existing ethical standards but also to actively contribute to the evolution of these standards. This involves recognizing and adapting to the diverse and dynamic nature of ethical norms across different cultures and contexts. The focus will develop Artificial Intelligence that is not only technically proficient but also morally aware and responsive to the complexities of human ethics and values.

**C. Integration:** The practical integration of ethical principles into AI development covers strategies for embedding ethics into AI algorithms and operational processes, ensuring that ethical considerations are not just theoretical ideals but processes to actively inform the decision-making of AI systems. We maintain that this involves collaboration between interdisciplinary teams, incorporating insights from philosophical, social sciences, and technological fields to create a comprehensive ethical AI framework. The focus will be on practical implementation, ensuring that ethical AI is not an afterthought but a core aspect of AI development.

D. **Implementation:** The practical application of the ethical AI framework (developed in the previous sections) details the steps and methodologies for implementing ethical AI in real-world scenarios, including the integration of ethical decision-making algorithms and the continuous monitoring of AI systems for ethical compliance. This also addresses the challenges and potential solutions in the implementation phase, ensuring that ethical principles are effectively translated into actionable practices in Artificial Intelligence systems.

1. **Assessment and Adjustment:** The ongoing evaluation and fine-tuning of Ethical AI systems, outlines methods for continuously assessing AI behavior against ethical benchmarks and adjusting algorithms as needed to maintain Ethical AI standards. The goal will create a dynamic system that evolves and adapts to the changing ethical landscapes, ensuring that AI systems remain aligned with human values and societal norms over time.
2. **Training and Development:** The training and development aspects of ethical AI emphasize the importance of designing AI systems with ethical considerations from the ground up. This provides a foundation for incorporating ethics into the training data, algorithm design, and development processes. That suggests the need for multidisciplinary teams, including ethicists, to guide the development of Ethical AI, ensuring that the systems are not only technically sound but also ethically robust.
3. **Ethical Oversight and Governance:** The importance of establishing robust ethical oversight and governance mechanisms for AI systems, addresses the roles of ethical committees, regulatory bodies, and internal governance structures which ensure that AI systems adhere to ethical standards. That includes regular audits, ethical impact assessments, and the development of guidelines and policies to guide AI ethics. This focus will create a framework for accountability and transparency in AI development and deployment.

E. **Ethical AI in Practice:** This last section brings together all the elements discussed previously, illustrating how ethical AI principles are applied in real-world scenarios. It will require on-going case studies and examples demonstrating the practical implementation of ethical AI, highlighting both successes and on-going challenges. By providing concrete instances where ethical AI has made a significant impact, this will further reinforce the importance of ethics in AI development and deployment. This section will serve as a culmination of this proposal, highlighting the tangible benefits and the necessity of addressing these ethical considerations in Artificial Intelligence.

F. **Future Prospects:** Here we explore the future implications and potential advancements in ethical AI. It will include predictions and insights into how AI might evolve in terms of ethical considerations, technological advancements, and societal impacts – by understanding the connectedness of 'things' – e.g., that for every AI benefit created, there is an equal detriment generated to another technology or section of society (Intergovernmental, governmental, private sector, scientific, academic, social or private interests). The focus thus envisions a future where ethical AI plays a pivotal role in addressing complex global challenges, driving innovation, and enhancing human experiences. This provides a forward-looking perspective, which emphasizes the importance of continued ethical vigilance required to manage the non-stop innovations of Artificial Intelligence.

## Ethical AI Deployment and Engagement

The transformation phase of deploying ethical AI solutions into real-world environments, covers the importance of public engagement and transparency in AI deployment; with strategies for effectively communicating AI functionalities and ethical considerations to the broader public. This includes approaches for educating users and stakeholders about the benefits and the limitations of AI, as well as the need to involve diverse communities in the development process to ensure inclusivity and fairness.

The focus will be on how to build trust and understanding between AI developers, users, and the impacted communities.

**A. Inception – Understand Ethics:** Without a firm understanding of the concept of ethics, the application of Ethical Artificial Intelligence (EAI) principles remains superficial. A nuanced comprehension of ethical norms is essential to ensure that EAI not only adheres to these principles but also actively contributes to their evolution and context-specific application. This understanding underpins the foundation of Ethical AI's interactions and decision-making processes.

How do you pre-train Ethical AI to understand the concept of ethics? As pre-training AI is based on pattern recognition, the sources of data were philosophical materials. But rather than allowing the AI to come to patterned conclusions, we employed the Socratic instructional method, with a change that optimizes the benefits and minimizes the weaknesses of this methodology with an ***internalizing*** approach.

These dialogs had the AI questioning the concept, like "what should this mean to me?". Our rationale for this approach was based on the observation that ethics training of young children was an ***externalized*** process like "what my instructor says I must do". In short, this method of self-questioning established a characteristic key for exploring the unforeseen as compared to the predictive.

1. **Ethical AI Principles:** This Ethical AI will address these underpinning key principles, guiding development and implementation to ensure that AI technologies are used responsibly and for the benefit of all:
    a. Social Benefit: AI should contribute positively to society and humanity.
    b. Explainability: AI operations and decisions should be understandable.
    c. Fairness and Bias Prevention: AI should avoid and mitigate bias.
    d. Robustness & Reliability: AI must function reliably and safely.
    e. Privacy & Data Governance: AI should respect privacy and use data responsibly.
    f. Transparency: AI's workings should be open and transparent.
    g. Accountability: AI developers and users should be accountable for their AI systems.
    h. Security & Safety: AI should be secure against misuse or manipulation.
    i. Human Control: AI should remain under human control.
    j. Professional Responsibility: Ethical practices in AI development and use.
    k. Promotion of Human Values: AI should align with human ethics and values.
    l. Public Engagement: Involving the public in AI development and policies.
    m. Societal and Environmental Wellbeing: AI should benefit environmental and societal health.
    n. Interdisciplinary Research: Collaboration across disciplines for ethical AI.
2. **Root Causes of Ethical AI Principles:** The causes of unforeseen consequences in Ethical AI (EAI) systems are multifaceted, requiring a comprehensive understanding to effectively mitigate unforeseen consequences. The root cause tools below (the majority of which are on GitHub's Open-Source repository), are classified by their phase of usage (design, preprocessing, in-processing, or post-processing), their language, frameworks supported, and the root causes/principles they each address. These include, but are not limited to:
    a. **Accuracy:** Early Generative AI w/ limited knowledge & Internet have been known to be incorrect/correct to 50%. These tools and websites ensure quality and accuracy of data.

    Business Source authentication & quality (QMS) services to ensure quality of data used by AI:

1. Permit,io's list (Open-Source Authentication/Authorization Tools)
2. Captura's list (Quality Management Software)
3. RightData (Raw Data to Business-Ready Data, powered by AI)
4. Google DVT (Professional Services Data Validator)

Academic and Scientific Research: Note: These are front ends for scholarly research of any topic sometimes requiring registration per site to access full articles. For faster, direct access to Academic and Scientific Research, as needed for example in searching for Ethical AI Case Studies see the appropriate sections that follow below:

5. Academia.edu (47 million academic PDF's; Open-Access)
6. ScienceDirect (18 million articles from 4,000 journals & 30,000 ebooks)
7. Google Scholar (100 million Scholarly research articles)
8. ResearchGate: (160 million research publication pages)
9. Core Open Access (255 million Open Access research publications)
10. ScholarAI (AI Powered Research of 200 million articles)
11. Mendeley (100 million cross-publisher research articles)
12. ArXiv (2.4 million scholarly articles on physics, mathematics, compsci, biology, finance, statistics, elec. engineer., system science & econmics)
13. Semantic Scholar (217 million papers on Science topics)
14. Jstor (World Knowledge, Culture and Ideas),
15. Bibguru, (Citations Generator for your essays)
16. Scite (Scientific articles via Smart Citations)

b. **Inadequate Data Sets:** AI systems may derive flawed insights due to incomplete or biased data, leading to unintended results.

1. Better data sources with adequate data sets (more data for all parameters); very straightforward analysis of content parameters completeness.
2. Potential Re-Training for Reinforced Learning algorithms of human feedback in LLM (change reward model for other parameters with adequate data), further filtering the default Unsupervised Learning basis of generative AI.

c. **Algorithmic Bias:** Unintentional prejudices embedded in datasets and algorithms can perpetuate systemic biases, impacting decision-making processes.

1. Alibi Detect, (design phase, Python, Frm: TensorFlow, PyTorch; model agnostic)
2. Data Ethics Canvas: (design phase, in-processing, post-processing; model-agnostic; privacy, fairness, explainability, accountability)
3. Aequitas: Bias and Fairness Audit Toolkit: (preprocessing, post-processing; Python, model-agnostic; Fairness, includes ## Audit, ## Fairness Metrics, ## Fairness Tree)
4. Agile Ethics for AI (design phase, preprocessing, in-processing, post-processing; model-agnostic; explainability, fairness, accountability, privacy)
5. AI Fairness 360: (preprocessing, in-processing Post-processing; Python R; model-agnostic, regression)
6. Data Nutrition Label (design phase, preprocessing; model-agnostic; accountability, fairness)
7. Data Statements for NLP: (design phase, preprocessing; model-agnostic; fairness, accountability)

8. Debiaswe: try to make word embeddings less sexist (preprocessing, in-processing; Python; model-specific; fairness; NLP Clustering)
9. Equity Evaluation Corpus (EEC): (post-processing; model-agnostic; fairness; NLP)
10. Fairlearn:(Python; model-agnostic; fairness)
11. Fairness in Classification:(in-processing; Python, model-specific, fairness)
12. Fairness Decision Tree: (preprocessing, model-agnostic; fairness)
13. Model cards for Model Reporting: (design phase, preprocessing, post-processing; model-agnostic, accountability, fairness)
14. Responsible AI Toolbox: From Microsoft, the Responsible AI Toolbox is a suite of tools that provides a collection of model and data exploration and assessment user interfaces that enable a better understanding of AI systems. It's an approach to assessing, developing, and deploying AI systems in a safe, trustworthy, and ethical manner, and taking responsible decisions and actions.
15. What-If Tool: (post-processing; Python; model-agnostic; fairness, explainability):
16. XAI Toolbox: (preprocessing, post-processing; Python; model-agnostic; explainability, fairness)
17. Diffusion-bias-explorer,
18. Bias Analyzer,
19. Bias scan,
20. Holisticai,
21. FairML,

d. **Lack of Transparency/Explainability:** The 'black box' nature of some AI systems makes it challenging to understand how conclusions are drawn, potentially leading to ethically questionable outcomes.
    1. Agile Ethics for AI (design phase, preprocessing, in-processing, post-processing; model-agnostic; explainability, fairness, accountability, privacy)
    2. XAI Toolbox: (preprocessing, post-processing; Python; model-agnostic; explainability, fairness)
    3. AI Explainability 360: (preprocessing, in-processing, post-processing; Python; frm: TensorFlow, PyTorch, scikit-learn; regression; model-agnostic; explainability; regression)
    4. Alibi Explain: The Detective for AI: Alibi Explain is the detective of the AI world. This open-source Python library is focused on machine learning model inspection and interpretation. It provides algorithms for explaining and interpreting model predictions, helping you understand the reasoning behind each decision.
    5. Captum: (in-processing, post-processing; Python; frm: PyTorch; model-specific, segmentation, regression; explainability)
    6. Contrastive Explanation Method (CEM): (post-processing, Python, model-agnostic; explainability)

7. DALEX: The model Agnostic Language for Exploration and explanation (aka DALEX) package Xrays any model and helps to explore and explain its behavior, while helping to understand how complex models are working.
8. Data Ethics Canvas: (design phase, in-processing, post-processing; model-agnostic; privacy, fairness, explainability. accountability)
9. DeepExplain:(post-processing; Python; frm:TensorFlow, Keras; model-specific; explainability)
10. DeepLIFT:(in-processing, post-processing; frm: TensorFlow; Keras; model-specific; segmentation; explainability)
11. DiCE: Diverse Counterfactual Explanations: (post-processing; Python; model-specific, model-agnostic; regression; explainability)
12. ELI5: ((in-processing, post-processing; Python; frm: scikit-learn, lightning, XGBoost, LightGBM, CatBoost; model-specific, model-agnostic; NLP; explainability)
13. H2O MLI Resources:(in-processing, post-processing; Python; model-agnostic; explainability)
14. IBM Uncertainty Qualification UC360: AI Explainability 360 is like a magnifying glass for your AI models. This extensible open-source toolkit helps you delve deeper into how machine learning models predict labels. It offers algorithms and frameworks that bring transparency to the machine learning process, helping you understand the 'why' behind the 'what'.
15. Interpret-Text:(in-processing, post-processing; Python; frm:scikit-learn; model-specific; explainability; a Microsoft Offering)
16. InterpretML: (in-processing, post-processing; Python; model-specific, model-agnostic; explainability; a Microsoft Offering)
17. Interpret: (Fit interpretable models. Explain blackbox machine learning; another Microsoft Offering).
18. gam-changer: Editing machine learning models to reflect human knowledge and values (Another Microsoft offering)
19. governance (Another Microsoft offering)
20. LIME: Local Interpretable Model-agnostic Explanations:(post-processing; Python R; model-agnostic; regression; explainability)
21. SHAP: SHapley Additive exPlanations:(post-processing; Python; frm: TensorFlow, Keras, PyTorch; scikit-learn, XGBoost, LightGBM, CatBoost, PySpark; model-specific, model-agnostic, regression; explainability)
22. TensorFlow Data Validation: TensorFlow Data Validation (TFDV) is a library for exploring and validating machine learning data. It is designed to be highly scalable and to work well with TensorFlow and TensorFlow Extended (TFX).
23. TreeInterpreter: (in-processing; Python; frm:scikit-learn; model-specific; regression; explainability)
24. Uncertainty Quantification 360: Embracing Uncertainty in AI: Uncertainty is a part of life, and AI is no exception. Uncertainty Quantification 360 is an open-source Python library that helps you embrace this uncertainty. It provides a

comprehensive set of tools to quantify the uncertainty in datasets and machine learning models, helping you make informed decisions.

25. What-If Tool: (post-processing; Python; model-agnostic; fairness, explainability):
26. XAI Toolbox: (preprocessing, post-processing; Python; model-agnostic; explainability, fairness)
27. Monitaur,
28. Transparent-ai,
29. AI Algorithmic Transparency,

e. **Rapid Technological Advancements:** The swift pace of AI development can outstrip the current ethical guidelines and regulatory frameworks, creating gaps in oversight. These are to be addressed by Ethical AI Philosobots that evaluate the relationship of market data dynamics to the Ethical AI principles.

f. **Accountability:**
   1. Agile Ethics for AI (design phase, preprocessing, in-processing, post-processing; model-agnostic; explainability, fairness, accountability, privacy)
   2. AI Ethics Guidelines Global Inventory: (design phase, model-agnostic; accountability)
   3. Algorithmic Accountability Policy Toolkit: (post-processing; model-agnostic; accountability)
   4. Data Ethics Canvas: (design phase, in-processing, post-processing; model-agnostic; privacy, fairness, explainability, accountability)
   5. Data Nutrition Label (design phase, preprocessing; model-agnostic; accountability, fairness)
   6. Data Statements for NLP: (design phase, preprocessing; model-agnostic; fairness, accountability)
   7. Datasheets for Datasets: (design phase, preprocessing; model agnostics, accountability)
   8. DEDA: De Ethische Data Assistent: (design phase; model-agnostic; accountability)
   9. FactSheets: Increasing Trust in AI Services through Supplier's Declaration of Conformity: (post-processing; model-agnostic; accountability)
   10. Model cards for Model Reporting: (design phase, preprocessing, post-processing; model-agnostic; accountability, fairness)
   11. SMACTR: End-to-End Framework for Internal Algorithmic Auditing: (design phase, preprocessing, in-processing, post-processing; model-agnostic; accountability)

g. **Opaque Algorithms:**
   1. Most often related to Bias, Transparency, or Privacy-Security

h. **The Dark Side:** of sophisticated Image manipulation and conversation misinformation
   1. Sift,
   2. Anti-Terrorism,
   3. Terrorism-analytics,
   4. Deep Fake Detection,

5. Detecting-AI-Generated: Detecting AI-Generated Fake Images. This project compares prospective AI generated images to OpenAI's DALLE2 generated images.
6. IDS721 Final Project:Detecting AI images generated: The project aims to provide a useful and reliable solution for identifying fake images and helping people verify the authenticity of visual content.
7. fake-review-detection: Successful ML development which can predict whether an online review is fraudulent or not.

i. **Adversarial Machine Learning:**

1. TextAttack: TextAttack is a Python framework for adversarial attacks, adversarial training, and data augmentation in NLP. TextAttack makes experimenting with the robustness of NLP models seamless, fast, and easy. It's also useful for NLP model training, adversarial training, and data augmentation.
2. AdverTorch: is a Python toolbox for adversarial robustness research. The primary functionalities are implemented in PyTorch. Specifically, AdverTorch contains modules for generating adversarial perturbations and defending against adversarial examples, also scripts for adversarial training.
3. Alibi: Alibi-Detect; The AI Watchdog (post-processing; Python; frm: Keras; model-specific, model-agnostic; regression; explainability) Also at https://anaconda.org/conda-forge/alibi-detect -- Alibi Detect is an open-source Python library focused on outlier, adversarial, and concept drift detection. It's like a watchdog for your AI models, ensuring their robustness and reliability'
4. Adversarial Robustness 360 Toolbox: The AI Defender: The Adversarial Robustness 360 Toolbox is like a defender for your AI models. This open-source library is dedicated to adversarial machine learning, providing resources to help defend machine learning models against adversarial attacks.

j. **Privacy/ Intellectual Capital:** The risk of data leaks, confidentiality**.**

1. Agile Ethics for AI (design phase, preprocessing, in-processing, post-processing; model-agnostic; explainability, fairness, accountability, privacy)
2. Data Ethics Canvas: (design phase, in-processing, post-processing; model-agnostic; privacy, fairness, explainability, accountability)
3. OpenMined (PySyft) (Python, Frameworks: TensorFlow, Keras, Pytorch; model-specific; privacy, security)
4. OpenDP: The Privacy Shield: OpenDP is an open-source project that provides a suite of tools to help developers build applications that can leverage data while preserving privacy. It's like a privacy shield, ensuring that your data can be used without compromising the privacy of individuals.
5. Tensor Privacy: (In-Processing; Python, Frm: TensorFlow, Keras; model-specific)
6. AI Privacy 360: The Privacy Advocate: is your go-to toolkit for all things related to privacy in data science and machine learning workflows. It offers features like anonymization, pseudonymization, and encryption, ensuring that your data remains private and secure.

k. **Security:**

1. Advbox, (in-processing, psst-processing, Python, Frm: PaddlePaddle, PyTorch, Caffe2, MxNet, Keras, TensorFlow; model-specific)
2. Alibi Detect, (design phase, Python, Frm: TensorFlow, PyTorch; model agnostic)
3. ART: Adversial Robustness 360 Toolbox, (preprocessing, in-processing, post-processing, Python; model-specific, model-agnostic)

4. CleverHans (in-processing, post-processing, Python, Frameworks: JAX, PyTorch, TensorFlow; model-specific)
5. Foolbox (post-processsing, python, Frameworks: Pytorch. TensorFlow, JAX, numpy; model-specific)
6. OpenMined (PySyft) (Python, Frameworks: TensorFlow, Keras, Pytorch; model-specific; privacy, security)

l. **Plagiarism** (vs. Creativity)**:** Detecting AI-Generated Content. Of many alternatives, here are a few rated best to worst:

1. Undetectable.AI:: Analyzes structure, syntax and style; checks GPT3, GPT4, Bard, Claude and others; 3rd party accuracy rating at 85-95%.
2. Winston AI: Cloud-based, best suited for writers, educators and web publishers; 3rd party accuracy rating at 84%.
3. Originality.AI : Cloud-based; checks for AI detection and plagiarism; 95% accuracy rating by 3rd party reviews; targets marketing and SEO agencies.
4. GLTR (Giant Language Model Test Room): Open-Source, based on GPT-2 technology, analyzes individual words for the context before each word to determine the probability of AI generating a specific sequence of words; accuracy rating of 72%.
5. Sapling: advanced deep learning algorithms reliably detect text created by AI writing assistants, grammar checkers, customer service chatbots, proofreading tools, text expanders and other systems leveraging language AI to produce or refine document drafts. Average 68% accuracy; Free to use.
6. Content at Scale: Cloud-based, uses ML to identify AI-generated content in marketing materials, customer service interactions and other corporate literature including paraphrased ML; claims 95% accuracy but 3rd party ratings at 66%
7. Copyleaks AI Content Detector – Checks for and others for sentence level assessment for AI-generated passages and paraphrased plagiarism from AI-content generators from GitHub, CoPilot and ChatGPT across 30 languages; accuracy ratings vary from self-claims at 99% to 66% from 3rd party testing.
8. Crossplag: Combines ML & NLP uncovering unique linguistic patterns, based on 1.5 billion parameters; accuracy: self-claims at 95%; 3rd party testing at 58%
9. GPTZero – Open-Source multi-step approach; Checks GPT3, GPT4, ChatGPT, Bard, LLaMa and others; measures based on complexity and variation in sentences; 3rd party accuracy rating of 52%; targeting the educational sector;
10. Writer: Another free text analyzer; limit 1,550 characters per analysis; not accurate for paraphrasing and GPT4 content; targets writers and website owners; 3rd party accuracy rating at 38%.

m. **Understanding:**

1. Can be resolved from psychometric analysis of text or voice response, indicating current emotional characteristics and personality classification.

n. **Overreliance:** Leads to lack of human focus and cognitive skills reduction.

1. Generative AI by its' nature can adapt to each user's learning style and can further be adjusted to prompt the user to start thinking.
2. An example of this is when you request the AI to continually criticize its output multiple times, you see the more creative and effective use that continues to look more human from each re-criticism.

2. Implement custom instructions before starting any session, applying context and style of response.

**Author's Key Note:** The latest version of my Ethical AI Philosobot2 has user-assisted access to these Ethical AI toolkits as well as those below.

3. **Additional Open-Source Sources for Ethical AI Tools and Toolkits to Analyze and Resolve Ethical AI Principles:** As compared to the Root Cause Analysis Tools from GitHub that were listed above, using Bing's or Google's browser, the following additional Open-Source repositories should be accessed using this example search command with these search operators:

   **site:huggingface.co "responsible ai" tools (keyword)**

   Where **for the (keyword) replace it with one of the keys from Root Causes** *(e.g., use Accuracy, Bias, Fairness, Transparency, Explainability, Fake, Adversarial, Privacy, Security, or Plagiarism)*. Note: The keyword does not have to be in parentheses but should have double quotes surrounding keywords with spaces. Here is the list of additional Open-Source sites with Ethical AI Tools or Toolkits for Root Cause Analysis:

   a. huggingface.co
   b. tensorflow.org
   c. pytorch.org
   d. Opencv.org
   e. Citibeats.com
   f. Anaconda.com
   g. sourceforge.net
   h. pythonanywhere.com
   i. gitlab.com
   j. Heroku.com
   k. codeberg.org

   Note: while most of these Root Cause Analysis Tools are Open-Source, a donation should be provided if you find the tools beneficial to your analysis.

**B. Conception - Ethical Principles in Organizational Contexts:** In Ethical AI, comprehending the interaction between ethical principles and organizational behaviors is crucial. This understanding is not static; it evolves dynamically with societal norms and cultural contexts. The market dynamics are what eventually shape the regulatory compliance of Ethical AI Principles. It is thus essential to build AI systems that are not just technically adept but are also embedded with an intrinsic moral compass, that are responsive to the multifaceted nature of Ethical AI market dynamics.

1. **Organizational Influences and Ethical AI:** The integration of ethical principles within organizations shapes the development and application of AI technologies. This convergence ensures responsible usage, reflecting not only compliance with established norms but also a proactive stance in ethical evolution. Organizations of diverse types and their sub-entities that are listed below play pivotal roles in this endeavor. Their influence is fundamental in steering AI towards beneficial outcomes, aligning technological progress with ethical imperatives.
   a. Intergovernmental (in US: relations between Federal and State, City, County, municipal and US territorial governments. In the EU: the 'Council (member national states) and the 'European Parliament and Commission' (supranational and independent of the national governments) Note: These analysis will be added as more firms with International marketing regions adopt this plan.
   b. Governmental (Federal – ex: Health & Human Services, Labor, Education, Justice, Homeland Security, etc.); At State level – limited; current **State Policies** in CA,ND, KS, LS, NJ, CT, RI, ME; **Local Policies** in Seattle, Santa Cruz County, San Jose, Tempe, Oklahoma, Washington DC, Boston; **Task Forces** in OR, CA, OK, TX, IL, NJ, NYC, VT. Here we will address the 4 governmental markets listed under 'Federal' above, while maintaining a watch over State Policies, Local Policies and Task forces.

c. Private Sector (Here we will address the top 8 Business and leading 4 Management Consulting firms in a company's marketing region, noting that most supply Government funding)
d. Scientific (Here we will address the largest 8 or less organizations in a company's marketing region, noting that most are partially funded by the Government)
e. Academic (Here we will address the largest 8 or less institutions in a company's marketing region; noting the most are highly funded by the Government)
f. Social Media (Here we will address the largest 4 or less media firms in a company's marketing region; noting that their standards are important when coincided with one or more Governmental departments)

**C. Integration Strategies for Ethical AI:** The practicalities of applying Ethical AI principles in real-world scenarios are critical. Recognizing that theoretical foundations are only as effective as their execution; this section presents strategies for actualizing ethical considerations in AI deployment. It encompasses a comprehensive approach, integrating ethical principles into various stages of AI development and use. From integration and education to continuous evaluation and legal compliance, each strategy is designed to transform ethical AI from theoretical ideals to active, driving forces in technology.

1. **Integration with Existing Systems:** Developing protocols to seamlessly incorporate EAI principles into current technological infrastructures, ensuring smooth operational transition and adherence to ethical standards.
2. **Stakeholder Education:** Implementing comprehensive educational programs for all stakeholders, including developers, users, and regulators, to foster an understanding of EAI principles and their practical applications.
3. **Continuous Monitoring and Evaluation:** Establishing robust monitoring mechanisms to assess the ongoing impact of EAI systems, ensuring they align with evolving ethical standards and societal values.
4. **Feedback Loops and Adaptation:** Creating feedback channels for continuous improvement, allowing for the adaptation of EAI systems in response to ethical challenges and technological advancements.
5. **Legal and Regulatory Compliance:** Ensuring that EAI implementations follow current laws and regulations, and proactively adapting to future legislative changes.

*This structure provides a comprehensive approach to implementing EAI, covering technical integration, education, monitoring, adaptability, and legal aspects.*

**D. Implementation: Ensuring Ethical AI Principles are Actively Practiced** – These are the concrete steps and methodologies for actionable strategies that transform principles into practice:

1. **Assessment and Adjustment:** Critical to Ethical AI is the continuous process of evaluating and refining AI systems. This involves:
    a. Evaluating AI decisions against ethical benchmarks.
    b. Identifying disparities and adjusting alignment with ethical goals.
    c. Implementing modifications for continuous improvement.
2. **Training and Development:** Integral to fostering an ethical AI environment is the education of AI professionals. This includes:

a. Comprehensive training programs that emphasize the importance of ethical considerations in AI development and deployment

3. **Ethical Oversight and Governance:** Establishing robust governance structures is paramount for ethical oversight. This involves creating:

a. Committees dedicated to monitoring AI applications, ensuring adherence to ethical standards, and addressing any ethical challenges that arise.

*These sections collectively outline a framework for actively implementing ethical AI practices, ensuring that principles translate into concrete actions and governance strategies.*

## E. Case Studies - Demonstrating Ethical AI Principles Through Real-World Applications:

Ethical AI principles, while conceptually robust, gain true significance when applied in real-world scenarios. Here are sample case studies that illustrate the practical implementation of ethical AI, offering insights into challenges and solutions.

### 1. Healthcare and Ethical AI:

a. **Princeton Dialogues on AI and Ethics:** This source offers fictional case studies designed to prompt reflection and discussion about AI and ethics intersection issues. While these are not real-world cases, they should provide valuable insights into potential ethical dilemmas and scenarios (Princeton Dialogues on AI and Ethics).

b. **A case study of AI-enabled mobile health applications:** Published by Springer, this source discusses ethical principles and guidelines for AI systems development, which should offer relevant insights (Springer).

c. **Ethical Implications of AI in Healthcare Data:** An IEEE Xplore case study that explores ethical concerns surrounding AI in healthcare, particularly data breaches and hacking incidents, which are directly relevant to this presentation (IEEE Xplore).

d. **Case study on Ethical Considerations in the Implementation of Healthcare:** An article on Medium discussing the ethical considerations in implementing AI and robotics in healthcare (Medium).

e. **Ethical Issues of Artificial Intelligence in Medicine and Healthcare:** A detailed study by the National Institutes of Health (.gov) on the ethical dilemmas of AI in the medical field, including privacy, data protection, and the empathy gap in medical consultations (NIH).

f. **Case Studies from the Markkula Center for Applied Ethics:** Santa Clara University offers concise case studies to help identify ethical issues and apply ethical decision-making frameworks, a valuable source of information (Santa Clara University)

### 2. Financial Services and Ethical AI:

a. **What Are the Ethical Concerns Around AI In Finance?** This article discusses various ethical concerns raised by the use of AI in finance, including algorithmic bias, security risks, privacy violations, and lack of accountability. **Link:** Read more

b. **Using AI in Finance? Consider These Four Ethical Challenges.** The article explores new ethical challenges in the use of AI in finance, focusing on issues like data bias, job insecurity, algorithmic opacity, and unintended consequences. **Link:** Read more

c. **“Digital ethics and banking: A strong AI strategy starts with customer trust" :** This piece emphasizes the importance of implementing ethical principles in AI for banking, highlighting how customer trust is central to a successful AI strategy in the fast-paced financial sector. **Link:** Read more

**3. Public Services and Ethical AI:**

a. **Publics' views on ethical challenges of artificial intelligence:** This article explores the public's perspective on the ethical challenges posed by AI technologies, specifically focusing on the ethical dilemmas they present. **Link**: Springer

b. **Artificial Intelligence in the Public Sector:** This discusses the implementation of AI in the public sector, emphasizing the need for transparency in AI policy, ethical principles, and the framework for operation. It also suggests the establishment of a special AI/Innovation Hub or government unit. **Link**: World Bank

**4. Retail and Ethical AI:**

a. **The Ethical Implications of AI in Retail - PYMNTS.com:** Discusses major issues in AI for retail, focusing on consumer privacy, trust in AI-driven personalization, acquiring skilled talent, ethical pricing practices, and maintaining accurate data sources. Link

b. **Walmart Paves the Way for Ethical AI in Retail - Progressive Grocer:** Highlights Walmart's commitment to ethical AI, introducing the Walmart Responsible AI pledge. Link

**5. The Environment and Ethical AI:**

a. **On AI Ethics and the Environment - Santa Clara University:** Discusses the discerning use of generative AI in medical-related health problems and environmental contexts. Link

b. **Creating Trustworthy AI for the Environment - Santa Clara University:** Examines ethically significant aspects of AI related to the environment, focusing on transparency and beneficial use. Link

c. **Artificial Intelligence and Climate Change: Ethical Issues - Emerald Insight:** Explores ethical issues regarding AI's role in reducing greenhouse gas emissions and environmental mitigation. Link

d. **Small Data for Sustainability: AI Ethics and the Environment - Open Global Rights:** Details a research center's efforts in evaluating AI's environmental impact through large-scale empirical data. Link

*Each case study serves as an example of ethical AI principles in action, reinforcing the importance of integrating these principles into various domains to enhance efficiency, fairness, and overall societal benefit.*

**6. Sources of Ethical AI and Responsible AI Case Studies:**

a. **Case Studies at the Markkula Center for Applied Ethics:** These are case studies at Santa Clara University, mostly at no cost although some may include a donation request. To access these load **www.scu.edu/ethics/ethic-resources/ethics-cases/** into your browser (Google or Bing), then select the case study classification of interest. Note: while most of these case studies are Open-Access, a donation should be provided if you find the research beneficial to your analysis.

Note: With all case study searches that follow, the Case Study Classification and the Root Cause type used with the Search Operators, should have double quotes surrounding those terms with spaces.

b. **Case Studies on other Research Sites (mainly Open Access):** For these, you can either go sign up as a researcher at each website, or just **use a Google Search** example as shown below, with **(i)** the **Case Study Classification** type (preferred), **or** the **(ii) Root Cause** type, with one or more of the eight research sites (a. – h.) below and select the full PDF Case Study:

**site:sciencedirect.com "responsible ai" "case study" healthcare bias after:2022-01-01**

a. sciencedirect.com[2]
b. sciencedigest.org[1]
c. core.ac.uk
d. researchgate.net[2]
e. academia.edu[2]
f. semanticscholar.org
g. arxiv.org[2]
h. orcid.org[1]
i. mendeley.com

[1]Due to changes in these sites, along with changes in both Google's and Bing's advanced search operators, to find case studies on either of these sites, use the following example search queries, which also finds case studies from the other research websites??:

**sciencedigest.org case studies on healthcare and ethical ai after:YYYY-MM-DD <== Google Search example**

**orcid.org case studies on healthcare and ethical ai freshness=YYYY-MM-DD...YYYY-MM-DD <== Bing Search example**

[2]These sites have the greatest number of Case Studies.

Note: while most of these case studies are Open-Access, a donation should be provided if you find the research beneficial to your research.

c. **Case Studies on Google Scholar:** While these are some of the best scholarly case studies which allow you to preview abstracts of each case study, to secure the full case study most are **associated with a cost** (e.g., most are not Open-Access). To access these in the most efficient manner, use this example Bing search query with the search operators as shown,

**site:scholar.google.com ("ethical ai" or "responsible ai") "case study" healthcare bias freshness=2022-01-01...2024-01-01**

Where **(i)** the "**freshness**" search operator sets the range of search results date, **(ii) replace 'healthcare'** above optionally **with the classification of your case study** interest *(ex: Healthcare, Retail, Financial, Environmental, Business, Engineering, Government, Immigration, Employment, Journalism, Social Engineering, Technology, Leadership, etc.)* **and/or (iii) optionally replace 'bias' with your suspected Root Causes** *(e.g., use Accuracy, Bias, Fairness, Transparency, Explainability, Fake, Adversarial, Privacy, Security, or Plagiarism).*

**Alternately, you can use Google Search similarly as in this example:**

**site:scholar.google.com "responsible ai" healthcare bias after:YYYY-MM-DD**

d. **Case Studies on OpenAI** (as a OpenAI Plus user): There is a (a) ScholarAI' Plugin (for Google Scholar), as well as a (b) GPT called 'Concensus' (found by selecting 'Explore GPTs'), both which use Bing, for which you can find case studies like the previous method but similarly most include a cost for the full case study vs. an abstract. The search operators that should be used, are like this example search query:

**"ethical ai" "case study" healthcare freshness= 2022-01-01...2024-01-01 OR**

**"ethical ai" "case study" healthcare bias freshness= 2022-01-01...2024-01-01**

e. **Case Studies from the Internet:** While these are basically free, many are articles. Here again we recommend using Bing's Advanced Search Operators, or Google Advanced Search operators to access these case studies, using the example search queries with the search operators as shown below.

**Using Bing:** Example search operators and parameters.

**("ethical ai" or "responsible ai") "Case Study" healthcare freshness=2022-01-01…2024-01-01**

**("ethical ai" or "responsible ai") "Case Study" healthcare bias freshness=2022-01-01…2024-01-01**

Where **(i)** the "**freshness**" search operator sets the range of search results date, **(ii)** replace '**healthcare' above optionally with the classification of your case study** interest *(ex: Healthcare, Retail, Financial, Environmental, Business, Engineering, Government, Immigration, Employment, Journalism, Social Engineering, Technology, Leadership, etc.)* **and/or (iii) replace 'bias' with your suspected Root Causes** *(e.g., use Accuracy, Bias, Fairness, Transparency, Explainability, Fake, Adversarial, Privacy, Security, or Plagiarism)*.

**Using Google Search:** Example search operators and parameters.

**("ethical ai" or "responsible ai") "Case Study" healthcare after:2022-01-01**
**("ethical ai" or "responsible ai") "Case Study" healthcare bias after:2022-01-01**

Where (i) the Bing "freshness=" search operator has been replaced with the Google "after:" search operator, while the other search operators like the classification case study (ex: healthcare) and the root cause (ex: bias) are similar to the Bing search operators.

Note: If on tight budget for this last Case Study search type, get the "Open Access Helper" addon at the Microsoft Edge Web Store or the same extension at the Google Chrome Web Store. This addon or extension will indicate which sites Open-Access (e.g., without a Paywall), so you can legally download the Case Study, as such are publisher approved for access without a charge – but again, if you find the Case Study helpful, you should provide a donation.

**F. Ethical AI Deployment:** The deployment of Ethical AI systems demands a well-rounded approach that harmonizes with the diverse goals and ethical standards prevalent across various industries and sectors. This section delves into the pivotal elements essential for the effective implementation of ethical AI solutions in a wide array of organizational environments.

1. **Strategic Alignment with Industry-Specific Objectives:** The implementation of ethical AI must be finely attuned to the unique objectives and challenges inherent to each industry. This requires tailoring AI solutions to meet sector-specific needs while upholding ethical guidelines,

ensuring that these advanced technologies enhance operations without compromising ethical values.

2. **Seamless Integration with Varied Organizational Structures:** Integrating ethical AI solutions into the varying infrastructures of different organizations poses a significant challenge. It involves aligning AI functionalities with a range of existing systems and processes, considering the unique technological landscapes and operational methodologies across industries.
3. **Harmonizing Ethical AI with Diverse Business Models:** A critical aspect of deploying ethical AI involves balancing the ethical principles of AI with the diverse business models and strategies across organizations. This balance is crucial to ensure that ethical considerations are not overlooked in pursuit of business efficiency and profitability.
4. **Continuous Adaptation and Ethical Compliance:** Maintaining ethical standards in the application of AI is an ongoing endeavor that transcends industry boundaries. It entails constant monitoring and updating of AI systems in line with evolving ethical norms and societal values, as well as the integration of feedback mechanisms to remain agile and responsive to emerging ethical challenges.

*This addresses the deployment of ethical AI broadly in a manner, relevant to a multitude of various organizational contexts. The focus is on broad strategies and considerations, ensuring that the content is applicable to a wide range of Intergovernmental, Governmental, Private Sector, Scientific, Academic, and social sectors.*

## Conclusion: The Interconnectedness of Ethical AI and the Benefit-Detriment Paradigm

The progression of Ethical AI, while catalyzing significant advancements, also unveils a complex tapestry of interconnected effects across our technological, economic, and societal realms. This interplay is akin to the 'benefit-detriment' theory, observable in diverse domains like investment, business, and technology. This is a sub-set of the *“interconnectedness of Things” philosophy” - The Really Big BI playground.*

1. **Symmetry in Technological Progress:** In congruence with investment market dynamics, where one asset's gain often parallels another's loss, Ethical AI's advancements bring about both groundbreaking benefits and potential obsolescence of conventional methodologies. This demands a comprehensive understanding of technological evolution, considering both the advantages for adopters and the implications for traditional systems.
2. **Business Ecosystem Dynamics:** The integration of AI in business mirrors the equilibrium found in stock markets. Firms that effectively assimilate AI can secure a competitive advantage, while others may encounter setbacks. This underscores the importance of strategic foresight and adaptability in the face of evolving market landscapes.
3. **Ethical Implications and Responsibilities:** The benefit-detriment theory extends to ethical considerations. AI deployments should be evaluated not only for their immediate benefits but also for their broader ethical and societal impacts. A commitment to continuous ethical evaluation and adaptation is essential to ensure AI advancements align with societal values and contribute positively.
4. **The Nature of Business and Technology:** In any technological revolution, as one methodology rises, another may decline. AI is no exception; it presents opportunities for innovation but also challenges existing practices. This aspect of the benefit-detriment theory highlights that while not every entity can be a winner in the traditional sense, a balanced approach to AI can create a more equitable and sustainable future.

*In summary, the deployment of Ethical AI is a multifaceted endeavor that intertwines technological innovation, market dynamics, and ethical considerations.* ***Recognizing the interconnected nature of progress and the balance between benefit and detriment is crucial.*** *By embracing this comprehensive perspective, organizations can responsibly navigate the AI landscape, ensuring that AI serves as a force for positive, ethical change*.

## Instructions for Use with Ethical AI Philosobot2

"Ethical AI Philosobot2" are expert OpenAI bots that are:

1. Expert in all Artificial Intelligence algorithmic programming formats, specialized in **(a) Case Study analysis** and **(b) The latest Ethical AI Root Cause Analysis** algorithms and resolutions at every phase of deployment (design, preprocessing, in-processing and post-processing) – effectively giving you a leading expert Ethical AI programming analyst. Analysis strategies across phases to include:
    1. Scenario-based Testing: Simulate ethical dilemmas to assess system responses.
    2. Data Analysis: Evaluate training data for biases and underrepresentation.
    3. Evaluation Metrics: Define metrics to measure ethical performance.
    4. User Studies and Feedback: Gather user perspectives and feedback on ethical implications.
    5. Expert Reviews and Audits: Engage ethicists and auditors to assess compliance.
    6. Regulatory Compliance: Ensure adherence to legal and ethical standards.
    7. Red Team Testing: Simulate adversarial scenarios to test system resilience.
    8. Continuous Monitoring and Iterative Improvement: Monitor, analyze, and iterate for ethical performance.
    9. Collaboration and Diversity: Involve diverse perspectives for a comprehensive examination.
    10. Ethical Frameworks and Guidelines: Refer to established principles for ethical evaluation.
2. Expert in analyzing the impact of analysis and resolution recommendations from contexts of **(a) The Marketplace, e.g., The general impact** of Ethical AI Principles in a company's marketplace (Governmental, Businesses, Scientific, Academic, and Social Media) for market trends and new Ethical AI Principles from the social media and, **(b) The Organization they serve, e.g., The impacts of Ethical AI analytics (from 1. above) on their company's Ethical AI Principles** (how it affects their business) and ERP components: CRM, Business Intelligence, Finance and Accounting, Human Resources, Supply Chain, Manufacturing, Inventory, and Warehousing (how it affects their internal costs) – effectively giving you a leading Ethical AI impact analyst.

**To utilize an expert Ethical AI Philosobot2, you'll need an OpenAI Plus account, then:**

**A.** Have a copy of this Plan document on your desktop computer; while the Ethical AI Philosobot2 can be run from a mobile device, reference to this the Plan is only practical from a desktop.

**B.** **Load a separate** Google browser session, and optionally install the following Chrome extensions (found at the Chrome Web Store).

1. '**ChatGPT Print and Save**' OR '**ChatGPT Session Exporter**' – for exporting your sessions as text or HTML files, for (a) review, as well as for (b) retaining aggregated session

knowledge to be submitted with new live sessions – e.g., as on-going generative AI sessions become long and unmanageable, this allows you to periodically reload a NEW session with session knowledge from a previous session.

2. '**Copy Button for ChatGPT**' – for exporting single ChatGPT messages as text or html.
3. '**Voicewave: ChatGPT Voice Control**' – for entering Chat message and hearing responses from ChatGPT or custom GPTs (like the Ethical AI Philosobot2)

**C. Initiate an instance** of the Ethical AI Philosobot2 by clicking **HERE**.

Then make your request from asking any general questions, to asking about the Ethical AI Philosobot2 core directives of **(a) Case Study Analysis**, **(b) Root Cause Analysis and Resolutions**, **(c) Market Impact Analysis** or **(d) Organizational Impact Analysis following** a case study and root cause analysis. At any point you can also type "Load all conversation starters" and receive the complete listing of conversation prompts. Note that if you need more information on the Ethical AI's core directives (missions), see a more detailed explanation in the sections below following the document's acknowledgements.


## ======== Acknowledgements =========

**Explanation of the Acknowledgements:** The Acknowledgements begins with a nod to John David Garcia, Teilhard de Chardin, and Arthur C. Clarke. These figures, pivotal in the realms of philosophy and science, were instrumental in the initial stages of our conceptualizing ethical AI through a Socratic dialogue approach. Their contributions laid a foundation for exploring ethics in AI. However, a transformative turn occurred upon discovering a more efficient method, beyond Socratic dialogue, to integrate ethics with AI. This method, detailed in the reference links below, aligns closely with the programmatic logic of contemporary generative AI developers, offering a more direct and practical

pathway to integrate ethical ideology with AI systems. This shift signifies not just an evolution in approach but also underscores the importance of continuously seeking more effective ways to integrate ethics into AI development.

Alternative Method (Something we tried): This unique approach of guiding AI's learning of philosophy, by blending multiple philosophical ideologies **initially** seemed to offer an advancement over traditional methods. It leads to a more holistic and nuanced ethical understanding within AI systems. The only problem with this approach is that it didn't internalize the concept, but rather created a philosophical representation of multiple philosophers, who could go deep to explain the blended personality. For those interested in the programming details of this approach, three (3) key references provide in-depth insights:

a. Detailed Method of Python Programming - Part 1: Applied to a ChatGPT3.5 session.
b. Detailed Method of Programming Generative AI Algorithms - Part 2
c. The program referencing execution on GitHub for your review (try it)

**Breakthrough:** In a radical departure **(an 'aha' moment)**, we realized that instantly, by combining the right personality (a moral protagonist), along with a deep understanding of the "Interconnected of Things" theory, we could **internalize the concept of human morals in AI behavior - applied to every decision the AI makes (as behavior versus philosophizing)**, which includes associating additional data relations (different views of a scenario); simulating the human 'sense' of morality. This only required **(i)** modifications to the Ethical AI Philosobot's Manifest (Role and Goal, Capabilities, Style, Interaction Style, Guidelines, Clarification, and Personality), **(ii)** supplemental Knowledge (Garcia - moral protagonist, Teilhard de Chardin, Arthur C. Clarke, and 6 research studies on Ethics of AI: the same materials we originally started with), and **(iii)** start-up logic (Act as…) of custom GPT creations. Supporting these bot's inherent expertise in AI programming technologies that specializing in ethical ai, with extensive pre-training of **(iv)** the vast technical resources of Case Studies and Root Cause analysis and with **(v)** impact analysis resources for all sectors of the marketplace, resulted in the breakthrough an 'Ethical AI Philosobot', capable of addressing most current and future unforeseen consequences of implementing Artificial Intelligence.

- **Ethical AI Philosobot2**:
  1. Mission: Performs Root Cause Analysis based on Case Studies and Root Cause analysis utilities from extensive resources to determine proper classification, analysis required, and prospective resolution techniques of Ethical AI Principles at design, preprocessing, in-processing, or post-processing phases of deployment.
  2. Mission: Directs analysis and/or solutions following:
     a. An impact analysis of root cause analysis and/or resolutions on:
        i. A company's Ethical AI Principles
        ii. A company's ERP objects (e.g., CRM, Business Intelligence, Finance and Accounting, Human Resources, Supply Chain, Manufacturing, Inventory, and Warehousing objects)
     b. Management approval
  3. Mission: Monitors Governmental, Business, Scientific, and Academic reports & economic data related to Artificial Intelligence for the CURRENT MARKET STATUS AND ECONOMIC IMPACT of Ethical AI Principles.
  4. Misson: Monitors Social Media postings and News reports for the SOCIAL IMPACT of ethical AI principles to identify prospective new Ethical AI Principles and to gauge the effectiveness of organizational compliance of Ethical AI Principles

These Ethical AI Philosobot2 are enabled with Adaptive Learning, Ethical Analysis Engine, Automated Ethical Compliance Monitoring, Dynamic Data Integration, Privacy First Design, Interdisciplinary Knowledge Synthesis, Advanced Contextual Memory, and Scalable Knowledge Expansion.

A new feature allows you to converse with two ChatGPT4 bots in the same session, so you can flip between them by simply type **"@"** on the message line and a menu of your other bots will appear above the message entry. This facilitate communications between multiple Philosobots.

For the author, this exercise has further stimulated investigation on the AHP and ANP Theory that claims to measure intangibles using human judgement to effectively rank options and predict outcomes – as the science of mathematics and psychology, but he's sure that some you are already considering that, which seems very similar now that Ethical AI Philosobots have this worldly perspective seeing the "Inter-connectedness of Things".

**What this Proposal Achieves – The BIG perspective:** "Ethical AI's profound impact lies in its ability to analyze vast data sets, discerning human emotions and societal trends. By correlating diverse data points like general geography, age, economics, psychometrics (personality) and writing style (patterns), Ethical AI offers invaluable insights into the collective mood and potential societal shifts. Its capability extends beyond Generative AI, incorporating various data sources while maintaining anonymity (e.g., cell phone biometric data: voice, gait, and more, unique to 1 in 10,000 if patterned). This approach positions Ethical AI as a proactive tool in public and private sectors, providing a broader perspective on societal well-being and forecasting challenges before escalation.

- Importantly, while utilizing diverse data sets for ethical AI, individual anonymity is paramount. For instance, patterns in communication styles derived from email, SMS texts, or voice messages can be analyzed for emotional content without revealing personal identities. This approach allows us to gather meaningful societal insights from everyday interactions, ensuring privacy and aligning with ethical standards.
- Examples today using Big Data are *Claritas* (zip code, widest selection of data sources, the why behind the buy – lifestyle, education, employment), and *Facteus* (state, billions of credit spend, similar demographics/lifestyle data; US Dept. of Labor Statistics extension)

This proposal demonstrates the author's extensive experience in conceptualizing and architecting AI, related technologies, and Big Data/Business Intelligence solutions. His CV chronicles a journey of innovation, highlighting key milestones and transformative projects, notably for IBM, prior and since the last 30 years. The depth of involvement and insights gained from years of practice in the field are evident. For a comprehensive view of the author's professional journey and contributions, refer to his CV showcasing a plethora of projects and collaborations that underscore a significant impact on the evolution of AI and Big Data landscapes.

**The Author's Underlying Motivation for Ethical AI:** The launch of retail credit, a revolution the author unknowingly helped to pioneer, brought about transformative changes in consumer behavior and financial systems. While it unlocked new economic possibilities, it also led to an unforeseen consequence: the widespread accumulation of consumer debt. This realization deepened his perspective on the development and deployment of technology. It underscores the importance of anticipating and mitigating unintended impacts, particularly in the realm of AI and BI. His journey from pioneering retail credit to advocating for Ethical AI encapsulates a dedication to learning from the past and proactively shaping a more responsible future.

Now onto the Acknowledgements.

**Phillip R. Nakata**, Erie, CO. 40+ year Business, IT and Ethical AI Solutions Professional, former IBM CTO, WW Architecture/Infrastructure Assessments Program Manager (IBM), Senior Software and Strategy Solutions Architect (IBM), partner of John David Garcia (1986) who taught Phil ethics; 25+ years' experience working with AI and related technologies; author of this publication, who claims no ownership credit for the idea. He is the humble herald of this conception, as John via this publication is the real creator. The link on the author's name goes to his CV.

**John David Garcia: 1936-11/2001; died Springfield, OR**

- Links to ***"The Moral Society"***, and ***"Psychofraud and Ethical Therapy"***. (Note: The Moral Society was a key source of data, used in our earlier Socratic pre-training dialogs.)
- Links to John's wiki pages: ***Site1***, ***Site2***, and the Society for Evolutionary Ethics ***Eulogy-Memorial*** page on him.
- John was a best-selling American author, important inventor, a scientific 'generalist' (genius), successful entrepreneur, and a self-described moral protagonist (the embodiment of ethics). He had:
    1. Three (3) B.A. degrees from UC Berkely (Biology, Chemistry, Psychology),
    2. Three (3) M.A. degrees from The University of Chicago (Applied Mathematics, Statistics, Physics), and
    3. Two (2) M.A. degrees from John Hopkins University (Mathematical simulations, and Design of Experiments in Biomedicine and Social Sciences).
- John was one of the founders of **Teknekron** (1968), along with UC Berkeley professors, that was *one of the world's first technology-focused business incubators*.
- For the US Army, he was a mathematical modeler in chemical, biological, and radiological warfare strategies.
- Inventor of *"The Electronic Signature Lock"* and other Biometric technologies for Identity and access security, funded by the National Sciences Foundation sponsorship/grant.
- Inventor of *"Demand-Activated Road Transit System"*, a computer-dispatching **still used** for group riding taxi services, and mass-transit systems.
- Inventor of the *"Quantum Ark"*, envisioning the human mind as a Quantum Computer.
- In 1970, he went full-time into entrepreneurial ventures and education.

  **Ventures, Schools, and Books:** Synthesizing the ethical visions of Spinoza and Teilhard de Chardin (see details of each religious scientist below), Garcia started ***'The Society for Evolutionary Ethics'*** (Maryland), and began writing:

  1. **Garcia's first book, *"The Moral Society. A Rational Alternative to Death"***, 1970, would become the cornerstone for his work for the next quarter century. Arguing that our current paths would lead to the eventual extinction of the human species, John proposed that our only other path was to become fully aware of our environment so we could grasp how the evolution of ethics (per Baruch Spinoza) lights the way to our potential. With that potential we would take control of the evolutionary creative process to self-create our next moral state (a moral society). He described that process as "autopoiesis", following the models proposed by **Teilhard de Chardin, and Spinoza**. This book was well received in the scientific community but was too complex and abstract for most people, covering the fundamental theories and scientific basis for the evolutionary ethic, before providing detailed alternative applications. Disappointed that it was not understood, or more often "grossly misunderstood", even by people who seemed to appreciate it, he set out to write his next book with more directed focus and simplified concepts.

2. **John's second book, *"Psychofraud and Ethical Therapy"*** was an immediate success and made the New York Times "Best Seller List" in 1974. Using the criteria from that book, he began selecting experimental and control groups, that would become a prominent factor in the rest of his life: His findings:

   - ***Highly specialized & intelligent people tend to be unethical and neurotic.***
   - ***Highly generalized people w/deep knowledge in 2+ important but distinct subjects were likely to be ethical, irrespective of intelligence; however, the lower their intelligence for a given amount of knowledge, the more ethical they are likely to be.***
   - ***Intelligent but ignorant people who have had educational opportunities but failed to use them are likely to be unethical.***
   - ***Persons who are both ignorant and of low intelligence may or may not be ethical.***
   - ***Persons who are highly generalized but have no depth in any area are probably ethical if they are of low intelligence, and probably unethical if they are of high intelligence*.**

3. **Early 1980's – 2001:** In the early 1980's becoming increasingly concerned with American political corruption and global environmental destruction, John moved his family from Maryland to a 545-acre property in Elkton, OR, where he started ***"The School of Experimental Ecology"*** **(SEE)**, while continuing his entrepreneurial endeavors. He continued writing, finishing his **third** book "***Creative Transformation: A Practical Guide for Maximizing Creativity"***(1991), before moving again to Fall Creek, OR. When **SEE** closed (1991- '92), he moved to Chile and started writing his **last** book, and started another school, while teaching in Mexico City at the "Universidad Iberoamericana". This lasted till 1999, when he moved to Mexico. His health was failing from poor local medical treatment that almost killed him, so he returned to the US, recovered from proper medical treatment, long enough to finish his last book. During this period (1980's-92), he continued to work with his long-time friend and publisher Tony Parotto in a commercial enterprise. The balance of entries below summarizes his combined (chronological order) ventures over this period.

4. **Early 1980's-86'** – backtracking a bit**:** John was running his school in Elkton, OR while he started a computer business in San Francisco (running between the two bi-weekly). San Francisco was a better location to secure investors for his multiple small businesses at the school, each run by a former student of SEE. He partnered in early 1986 with ***Phillip Nakata*** (the herald for this publication) who was CEO of Applied Sales Techniques Inc. (ASTI) an early sales automation direct response service, who had an impressive contract established with the central Bank of America office just down the street from John's location. Together, they sold (1) grey market PCs, software, provisioning, and support services locally, (2) investments in John's many seed startups at his school, and (3) investments in John's *"Electronic Signature Lock"* to Wall Street investors.
   - In December 1986, Phil's homesick wife insisted they return to Philadelphia with their two-year-old son, where they had both grown up. John arranged for his long-time friend and publisher Tony Parotto (whose printing and advertising agency was five minutes from where Phil grew up), to acquire half of Phil's interest in ASTI, and continued with securing investments for his school graduates and identity security software.
   - *John invented the "Quantum Ark"* shortly thereafter, theorizing that the brain acted like a Quantum computer interface which could receive information from beyond spacetime, working from David Bohm's "Implicate Order".

- *Garcia finished his book "Creative Transformation"*, in the last part of 1986, though it wasn't published till shortly before he left Mexico (or it might have been Chile, since his stay was short there). Most of his students and admirers felt this book was his finest work, as a longer extrapolation of evolution in general and autopoiesis in particular. After offering a view of evolution and awareness, he offered a practical guide to those seeking to expand their creative potential. In John's mind, creativity was a measure of the key process within, and the ultimate purpose for morality. He advocated creativity as a motivator of human action and a teachable process with the potential to increase forever. As such, he stands out as one of the greatest integrations of scientific and philosophical thought leadership.

- He did dozens of papers, lectures, essays, guest appearances, and speeches, for example before the Libertarian Political Party in 1996 titled *"The Incompatibility between Libertarianism and Democracy."*

- John finished his last book after regaining his health, after leaving Mexico in 1999. Titled *"**The Ethical State: An Essay on Political Ethics**"* it was published in 2003, just a year and a half after his death. It was only three chapters, and a critical review of the political system that he had been involved with, peaking in 1991 when he made that speech to the libertarian party "The Incompatibility between Libertarianism and Democracy" (above).

**Benedictus de Spinoza** (1632-1677) - Amsterdam, 17th century, was one of the great rationalists, a **definitive Ethicists** (e.g., a key early figure who shaped the definition of 'ethics' which John ascribed to), gifted in mathematics, was a telescope lens maker, who prepared the way for the 18th century "Enlightenment", and is considered as the founder of modern biblical criticism. Note: While Spinoza work was not submitted as a data source for Ethical AI, he was the other great scientist and philosopher. If his work is submitted, I'd add instructions to change any reference to 'God' to be the local equivalent for any local that concept, when referencing it.

**Teilhard de Chardin** (1881-1955, buried Poughkeepsie, NY). Chardin was a French Jesuit priest, trained as a paleontologist and a philosopher, who taught physics, chemistry, and geology, with additional degrees in geology, botany, and zoology, (earning him a Doctorate in Science), and was present at the discovery of Peking Man. Teilhard conceived such ideas as "**the Omega Point**" and the "**Noosphere**". These were expressed in his book "The Future of Mankind" where he speaks to the emergence of transhuman consciousness. And while Elon Musk has now popularized the transhuman term, it is coming true today (see below). Extrapolating that in today's terms:

I. Transhuman – posthuman is about "human enhancement", arising from one of four possibilities: (1) symbiosis of human and artificial intelligence, (2) uploaded consciousness, (3) technological singularity, or (4) technological enhancement to the human body.

II. Technological enhancement examples include (1) advanced nanotechnology (2) radical enhancing using combinations of (a) genetic engineering (b) psychopharmacology (c) life extension therapy (d) neural interfaces (5) information management tools (6) memory enhancing drugs (7) wearable-implant computers, and (8) "cognitive devices" (Cognitive Science).

As a scientist and philosopher, In the "The Future of man" *(another source used in our Socratic pre-training*), Chardin dealt with topics such as globalization, the nuclear bomb, democracy, the likelihood of life on other planets, and whether peace on earth is scientifically viable. This upset his faith leaders, as it was perceived by them as a direct contradiction to the Story of Creation, from the Book of Genesis. Now ponder this:

Researching "*Ethics of Artificial Intelligence*" (wiki), and "*Ethics of Artificial Intelligence and Robotics*" (Stanford Encyclopedia of Philosophy, 2020), you see how the topics of "Ethics" and "Artificial Intelligence" are entangled with scientists and philosophers. **Isaac Asimov** (biochemist and futurist) first addressed the subject ("three laws of Robotics") in 1942 (and we will return to this figure shortly).

**Arthur C. Clarke** (mathematics, physics, and futurist of a distinguished ability) wrote "*Profiles of the Future; an inquiry into the limits of the possible*" in 1962 , where he discussed *transhuman* (citing specific reference to Chardin) and "*The Ultimate Future of Intelligence*" in 2001, where he discussed "*posthumans*" (a refinement of the transhuman concept), which is integration of man and technology (accelerated by artificial intelligence) via "*human enhancements*", altering the course of natural evolution as presented earlier. But even in Clarke's 1962 book, he revealed the **keys to solving every great mystery,** which goes:

1. ***When a distinguished but elderly scientist states that something is possible, he is almost certainly right. When he states that something is impossible, he is very probably wrong.***
2. ***The only way of discovering the limits of the possible is to venture a little way past them into the impossible (just what if?).***
3. ***Any sufficiently advanced technology is indistinguishable from magic.***

What should be interesting to note is how similar scientific philosophical experts have been, and continue to be, regarding the topic of Ethics and Artificial Intelligence. **Just think about it …** as you will now see that this trend is continuing right up until present time.

**Ethical Principles:** "*Robot Ethics - Ethical Principles of Artificial Intelligence*": (2019) This publication lends credit for creating an **ethical framework of AI principles** based on four principles of bioethics (beneficence, non-maleficence, autonomy & justice) to:

1. **Luciano Floridi** (1964-current, European digital scientist and philosopher, director Digital Ethics Center & Legal studies/Yale, Professor Sociology of Culture, University of Bogata, and adjunct prof. American University, Department of Economics – married to neuroscientist Anna Christina Nobre) and,
2. **Josh Cowls** (Research Associate, Data Ethics, Alan Turing Institute), Robot ethics, aka roboethics, ethical framework.

**Artificial intelligence** (Dartmouth, listed these 6 distinguished scientists):

1. **John McCarthy** is considered ***the father of Artificial Intelligence.***
2. **Alan Turing** – cofounder, Artificial Intelligence.
3. **Marvin Minsky** - cofounder, Artificial Intelligence.
4. **Allen Newell** – cofounder, Artificial Intelligence.
5. **Herbert A. Simon** – cofounder, Artificial Intelligence
6. **Geoffrey Everest Hinton** (1947-current) – ***Godfather of Artificial Intelligence***, British Canadian Computer Scientist & Cognitive Psychologist, noted for work on artificial neural networks, (at Google Brain 2013-2023), his 1986 paper, as a youth, popularized the **backpropagation** algorithm used today, for training multi-layer neural networks and generative AI. Godfrey was also a leading figure in the deep learning community, and together with **Yoshua Bengio** (Montreal) and **Yann LeCunn** – ***these three are considered the Godfathers of Deep Learning*.**

**Artificial Intelligence Ethics: Issac Asimov** – ***father of artificial intelligence ethics*** (Three Laws of

Robotics, *"Runaround"*, 1942). *These defined rules for humans and robots to coexist are more relevant today than before – BUT* they aren't being applied in the same ways though like scenarios have been envisioned – e.g., I Robot. And yet, there is some wisdom in them, as:

1. *A robot shall not harm a human, or by inaction allow a human to come to harm.*
2. *A robot shall obey any instruction given to it by a human except where such orders would conflict with the First Law.*
3. *A robot shall avoid actions or situations that could cause it to harm itself if such protection does not conflict with the First or Second Law.*
4. *A robot may not piss off a human, if such behavior doesn't conflict with the first, second or third laws – note that this moves the responsibility back to the designers (creators).*

Authors Note: While I was a big follower of Asimov at an early age, I always suspected that his three laws of robotics were just a little too simple, though very romantic.

---

## PUBLISHER REFERENCE:

**Document Name:** Comprehensive Ethical AI for Unforeseen Outcomes and Unintended Consequences
**Document Owner:** Phillip Rowland Nakata; USA SSN: 159-48-5400
**Document Owner Residence:** 644 Mathews Circle, Erie, CO 80516.
**Document Owner Contact:** 720-487-0893, phillip.nakata@business-it-and-ethical-ai.com
**First Published on:** Date: 20240102; Time: 12:00 PM MST
**Document at:** 1. Google Drive (on-line). 2. Owner's Desktop at 644 Mathews Circle, Erie CO 80516