Embracing Human-AI Partnerships by construction of a Modular neural network for Safe timely development of AI.

The question: What role should the government take in the Ongoing advancement of Safe Artificial Intelligence (AI)?

Author Ray Parker

ray@wijidij.com

0400667236

This paper was authored through queries to Chat GPT 3.5 and saved considerable hours through a human AI partnership. No humans were harmed during this process.

Executive Summary

Advancing AI Development through Collaborative Government Hosting

In the pursuit of safe and responsible AI development, Modular Neural Networks (MNN) have emerged as a promising AI architecture; Augmenting capabilities in Human-AI partnerships, mitigating risk profiles, while keeping an efficient open network. It is **proposed** that The Australian government, through the department of Industry, Science, and resources, develop its own MNN, with help and guidance from many industry experts and developers. In so doing, the department can meet the rapidly growing need for safe ethical AI development and fulfill obligations without a need for an ever-increasing resource. The launching of the platform will act as a catalyst to create a societal wide ecosystem that will promote trust in AI systems and through open collaboration be able to evolve processes to match accelerating developmental needs. This document outlines some beneficial aspects and steps to be taken to ensure the department can start simply while building a low-risk collaborative platform to assist and guide all aspects of safe AI development.

1. AI supported self-service applications will do the heavy lifting and help align different perspectives through common language and translation.
2. The platform will leverage industry interest, the need for compliance to regulation and a group of passionate advocates that have been collectively amassing a knowledge based suited to the needs of future growth.
3. Incorporation of Risk Analysis and Mitigation principles in alignment with AI capabilities can ensure a secure environment for stakeholders.
4. Collaboration between government bodies, industry partners, and research institutions can facilitate the exchange of knowledge, resources, and expertise, streamlining the AI development process.
5. Training programs and educational initiatives will be instrumental in nurturing AI literacy and fostering a culture of responsible AI usage.
6. As the deployment of Large Language Models and generative AI becomes more accessible and seamless, the opportunity to align human and AI capabilities will further augment outcomes.
7. Task displacement and resource realignment can be strategically managed to optimize productivity and efficiency.
8. The capacity to integrate AI systems into existing processes will pave the way for swift and efficient implementation.
9. Building a government-hosted AI platform will act as a catalyst for innovation and collective progress. By bringing together the expertise of diverse stakeholders.

10. The development process can be enriched, and AI applications tailored to address specific challenges faced by industries and organizations.
11. A collaborative approach will ensure that the platform is comprehensive, able to streamline regulatory compliance and be future proof.
12. Encouraging dialogue on AI ethics, safety, and best practices will inspire trust and public acceptance. Moreover, by adhering to basic safety principles and hazard analysis, the government-hosted AI will promote transparency, ethical conduct, and the responsible deployment of AI technologies.
13. Embracing collaborative efforts in developing a government-hosted AI platform is the gateway to a future of safe, ethical, and efficient AI applications.
14. By pooling resources, fostering partnerships, and investing in education, the government can lead the way towards unlocking the full potential of AI while ensuring its responsible integration within the society and industries.

The benefits of such a decision are vast, from improved safety and productivity to innovative solutions that positively impact the nation and the world at large. Please read the full document as there are many considerations that require simple first steps. I sincerely hope that this helps formulate decisions and would very much like to further consult with development. Unpredictability in AI development can be, if not foreseen, then catastrophic risks mitigated.

1. Introduction: The Proposal

AI has the capacity to affect any area of society collectively and influence every organisation or individual. All our technological systems contain data, that are derived from, or have been analysed by AI resources. Just like genetically modified crops, once released into the ecosystem, turning back is not an option. AI needs to be utilised now in all things to enhance and combat negative sociodynamic forces. Our knowledge will assist in the creation of safe, beneficial AI systems. Current development is typically controlled by shareholder needs and driven by profit motives. To foster safe ethical AI, government, and the diverse fabric of society, must create an ecosystem that enables controlled advancement, without stifling potential (Montes & Goertzel, 2019). With a collaborative ecosystem effecting change, the true benefits of AI can be harnessed for the good of all.

The Australian Government is the only societal organisation able to achieve, holistic coverage, and collaboration necessary, to keep AI safe and ethical(Chandler, 2013).

1.1. Government Hosted AI

For government to be effective, a controlled AI platform needs to be constructed for all levels of interest: user-guided, user-centric, all-inclusive. It should focus on the needs of the individual first and balanced against the needs of society. This will not be a technological leap and only requires specific application of existing, decades old, AI understanding. This platform will represent all frameworks, principles, guidance, support, and collaboration already applied but will provide a self-serve timely way of safe AI development with minimal resources in application.

1.2. High-Powered Computing (HPC)

With the advent of High-Powered Computing (HPC) theory, algorithms developed long ago have the potential to provide AI augmented growth, build trust for each person and the community.

High-powered computing has unlocked boundless potential for AI integration across society, revolutionizing industries, and everyday life. With immense processing capabilities, AI can now tackle complex challenges, optimize processes, and deliver innovative solutions, propelling us towards a future of endless possibilities.

1.3. AI development through collective human intelligence

For decades a branch of society has been collectively working together on the theory behind implemented AI and have forecast the potential while outlining a blueprint for success. This represents a collective

body of work, easily accessible by searching through academically published work. Beneficial change is being held back only by the fear of losing what we have (Montes & Goertzel, 2019).

We have technology where the needs of many, may not necessarily out way the needs of the few. Instead, the needs of all can be met, without detriment to any (Mintz, 2015).

1.4. Globalized Collaborative deployment.

By strategically applying safe AI in conjunction with IoT and robotics, a transformative shift in productivity, labour practices, and resource planning emerges. This powerful amalgamation promises to significantly boost Gross Domestic Product (GDP)

1.5. Platform essentials

A Platform, with immediate, ever-increasing capability to guide safe ethical AI advancement must be developed. It must, be accessible and connected to all, adapt to rapid change, embrace new tech, set vital standards, and meet fundamental needs (McLeod, 2018). This is the only, and best, option for Australia. Benefits, obstacles, Aspects of safety, implementation and AI development methodology are described as follows.

1.6. Platform architecture

Based on Modular Neural Network (MNN) architecture, devised in the 1980'yet very relevant to the future of AI. A MNN is a type of AI network architecture that is designed with modularity in mind. In this approach, the neural network is divided into distinct and relatively independent modules, each responsible for specific tasks or subproblems. These modules can be interconnected in various ways to achieve complex functionalities.

An MNN is designed to be plug and play, with common rudimentary code to help, link, add, evolve, or discard sections of the total MNN, without compromising integrity or efficiency. An MNN can be trained and worked on without disrupting effective use.

2. MNN Open architecture

2.1. Pros

2.1.1. Embracing open network architecture in AI is like opening floodgates of creativity and collaboration. Fostering a vibrant ecosystem where diverse minds come together to unlock the full potential of AI. With an open approach, researchers and developers can freely exchange ideas, methodologies, and breakthroughs, promoting a culture of innovation.

2.1.2. Transparency and accessibility are at the heart of open AI networks. By making code, data, and models openly available, we empower the broader community to understand, scrutinize, and improve upon existing algorithms. This empowers a sense of collective responsibility in creating AI systems that are unbiased, robust, and trustworthy.

2.1.3. Cooperation in open networks breeds cooperation among competitors. Collaborative efforts transcend individual ambitions, pooling resources, and expertise for grand challenges that would otherwise be insurmountable. Synergy emerges, taking AI to new heights, and accelerates progress in ways unforeseen within isolated silos.

2.2. Cons

2.2.1. Open network architecture comes the vulnerability of exploitation. Malicious actors can exploit the openness to manipulate models or inject harmful biases, raising concerns about security and ethics. As we embrace transparency, the MNN becomes a robust safeguard handling potential threats that as they arise with the immediate power of ever evolving algorithms.

2.2.2. The Openness can become a double-edged sword, exposing sensitive intellectual property (IP) and proprietary knowledge to unrestricted scrutiny. This may discourage some stakeholders from fully engaging in open collaborations, fearing the loss of competitive advantage or erosion of market share. This Platform must have safeguards for IP as well as capability to share ownership of ideas without fear of loss.

2.2.3. Navigating the ocean of contributions in open networks can be overwhelming. Without proper curation and validation, information may lead to confusion, misinformation, or replication of flawed findings. Rigorous peer review and validation mechanisms are essential to distinguish the gems from the noise but as the AI learns this task will need less resourcing and deliver higher accuracy.

2.3. Summation

An MNN should be open to user modification and guidance regardless of user coding knowledge, with an interactive large language Model (LLM) assistant and access to multiple data sources directly or indirectly in use. A generative model or Generative Adversarial Network (GAN) will also have value but is planned for next steps.

The open network architecture in AI breathes life into an exciting era of collective creativity and progress. It empowers us to build smarter, more inclusive AI systems, driven by transparency and community collaboration. Yet, we must tread carefully, guarding against potential misuses and ensuring that the benefits of openness outweigh its drawbacks. With diligence and prudence, open AI networks can usher in a golden age of artificial intelligence that serves humanity's best interests.

3. Security

3.1. Inherent security

MNN Architecture lends itself to common code that sorts, blocks the unwanted and delivers only the information required to satisfy user needs in a data exchange. This returns information ownership back to the individual, enhancing trust, reducing cost and potential liability/privacy concerns. Any MNN correctly constructed has innate security built in but is bolstered in ways that standard architecture is not. Blockchain methodology is inherent but can be applied to split packets and host them within across multiple data infrastructures without risk. This makes intrusions difficult, and patches easily added.

4. Information Exchange

4.1.1. The Platform's primary objective is to encompass an exhaustive database of relevant information, readily accessible to all stakeholders with vested interests. The information can either be directly incorporated within the Platform's database or accessed through interconnected links to reputable 3rd-party sources.

4.1.2. Over time, as 3rd-party entities begin to embrace and integrate their processes into the Platform, it will organically evolve into a cohesive, shared, and secure ecosystem. The evolution will bring about enhanced safety measures and ensure the Platform's integrity, fostering an environment of trust and collaboration among its users and contributors.

4.1.3. This evolution has precedence. It can by prediction, be correlated to the expansive growth of the Microsoft Windows operating system, which shaped personal computing worldwide (Moore & Tambini, 2018). However, the stakes are heightened in this context, as we centre our focus on meeting profound social needs, ensuring safety, and nurturing collective well-being. With these crucial factors at the forefront, the adoption of such transformative advancements is bound to be swifter, and the trust will build to a greater proportion.

4.2. Varied methods of Inputs/Outputs (I/O)

4.2.1. Modular neural networks possess remarkable capabilities in filtering and transferring data seamlessly between modules, enabling efficient communication and interaction with users. Through

this modular architecture, data can be processed more intelligently, as each module specializes in specific tasks, promoting optimized information flow and reducing redundant computations. This streamlined data exchange not only enhances the network's performance but also provides a more responsive and intuitive User Experience (UX).

4.2.2. The modularity of the neural network allows for effortless integration with various peripheral devices. By accommodating diverse I/O interfaces, the network can easily interact with sensors, cameras, microphones, and other external devices, thereby broadening its scope of applications. This versatility opens doors to novel use cases and more natural interactions with users, contributing to a more immersive and user-friendly experience.

4.2.3. Design facilitates the integration of generative AI capabilities seamlessly. As generative modules are added, the neural network can evolve dynamically, adapting to new tasks and generating synthetic data, images, or audio, among other possibilities. This adaptability enables the network to continuously improve and expand its functionality without the need for a complete overhaul, keeping pace with the latest advancements in AI.

4.2.4. MNN demonstrate their prowess by efficiently managing data flow between modules, optimizing interactions with users, and seamlessly integrating with peripherals. Their flexible architecture empowers them to evolve and embrace generative AI capabilities, making them powerful tools for enhancing User Experience and staying at the forefront of AI advancements.

5. First Steps for Government AI development

Developing a modular neural network linked to an online hosted Large Language Model (LLM) for the Australian government requires a systematic and well-structured approach. Here are the first steps to initiate the process:

5.1. Establish a Cross-Disciplinary Team:

Assemble a team comprising experts in AI, data science, cybersecurity, legal, and ethics to collaboratively design the architecture. This team should also include representatives from various industry sectors and relevant government agencies to ensure all stakeholders' perspectives are considered.

5.2. Define Objectives and Use Cases:

Clearly outline the objectives of the modular neural network and the online hosted LLM. Identify specific use cases where the network will be deployed to address industry-specific challenges or enhance organizational processes.

5.3. Data Collection and Curation:

Identify and collect high-quality and diverse datasets relevant to the use cases. Ensure that the data complies with privacy regulations and is representative of the Australian context. Curate the data to remove biases and ensure its suitability for training the neural network.

5.4. Selecting the Large Language Model:

Choose a well-established and reputable Large Language Model that aligns with the project's objectives and meets the necessary performance criteria.

5.5. Design Modular Architecture:

Develop a modular architecture that allows seamless data flow and interaction between various modules while ensuring scalability, flexibility, and ease of integration. Each module should serve a specific function, addressing different aspects of AI safety, such as bias detection, explainability, and robustness.

5.6. Safety and Ethics Protocols:

Incorporate safety and ethics protocols into the design to ensure responsible AI development. Establish guidelines for handling sensitive data, addressing bias, and maintaining transparency in the decision-making process.

5.7. Online Hosting and Security:

Select a reliable and secure online hosting service to deploy the Large Language Model. Implement robust security measures to safeguard data and protect the system from potential cyber threats.

5.8. Pilot Testing:

Conduct pilot testing of the modular neural network with a limited set of stakeholders to gather feedback, identify any issues, and fine-tune the system accordingly.

5.9. Educational Initiatives:

Promote educational initiatives to familiarize stakeholders with the modular neural network and the online LLM, ensuring proper understanding and responsible usage.

5.10. Deployment and Continuous Improvement:

Deploy the system for broader usage while continuously monitoring its performance. Regularly update the LLM and the modular neural network based on user feedback, technological advancements, and evolving safety standards.

5.11. First Steps Summation

By following these initial steps, the Australian government can lay a strong foundation for developing a cutting-edge modular neural network linked to an online hosted Large Language Model, empowering stakeholders across industries to develop safe and responsible AI solutions.

6. Next steps for government MNN development and expansion

Following the initial steps, the Australian government can proceed with the following next steps to further develop and optimize the modular neural network linked to the online hosted Large Language Model:

6.1. Scale-Up and Full Deployment:

Expand the deployment of the modular neural network and the online LLM to encompass a broader range of stakeholders across various industries and organizations. Monitor performance and gather feedback from a diverse user base to identify potential areas of improvement.

6.2. Collaborate with Industry Partners:

Collaborate with industry partners and research institutions to foster innovation and knowledge exchange. Encourage industry-specific customization of the modular neural network to cater to unique requirements and challenges in different sectors.

6.3. Research and Development:

Allocate resources for ongoing research and development to enhance the capabilities of the modular neural network and the LLM. Stay abreast of the latest advancements in AI safety and continually update the system to reflect the best practices.

6.4. Continuous Training and Fine-Tuning:

Regularly retrain the neural network using up-to-date datasets to keep the LLM relevant and accurate. Fine-tune the modular architecture based on real-world experiences and emerging challenges to ensure the system's adaptability.

6.5. Accessibility and Outreach:

Ensure that the platform is accessible to organizations of all sizes, including small and medium-sized enterprises. Implement user-friendly interfaces and provide comprehensive documentation and tutorials to encourage wider adoption.

6.6. Community Engagement and Feedback:

Engage with the AI community and encourage discussions on AI safety and ethics. Gather feedback from users, industry experts, and the public to drive continuous improvement and address emerging concerns.

6.7. Regulatory Framework:

Work with regulatory bodies to develop a comprehensive regulatory framework that guides the use of AI systems and ensures compliance with ethical standards. Foster an environment of trust and accountability for AI technologies.

6.8. Promote AI Literacy:

Launch public awareness campaigns to promote AI literacy and empower stakeholders to understand AI technologies better. Address misconceptions and demystify AI to build public confidence in AI systems.

6.9. Collaboration with International Partners:

Collaborate with international partners and participate in global discussions on AI safety and standards. Exchange knowledge and best practices to contribute to the responsible development of AI on a global scale.

6.10. Continuous Evaluation and Auditing:

Conduct regular evaluations and audits of the system to assess its safety performance and adherence to ethical guidelines. Transparently communicate the results of evaluations to the public and stakeholders.

By following these next steps, the Australian government can foster a thriving ecosystem of AI safety management, ensuring that the modular neural network and online hosted LLM are continuously improved, accessible, and aligned with the evolving needs and expectations of all stakeholders.

7. Safety and ethics

The platform is crafted to facilitate and expedite safe development in Australia. While the fundamental principle of "Do No Harm" underpins all rules, there exist more intricate methods to uphold and reinforce this ethical standard. Citizens rightfully expect more than mere safety assurance from a government funded primarily by taxpayers. This platform goes above and beyond to surpass expectations by adopting a people-centric approach, considering individual needs and perspectives. Leveraging advanced capabilities, such as the new generation of LLM have demonstrated. The platform can effectively cater to each person's unique requirements, even in small measures, contributing to coordinated and successful outcomes.

7.1. MNN architecture

is not Black boxed and should be open, as it is the best way for the AI to be trained, and the results for users immediate. However, this does not mean proprietary code contained in the platform should be shared with all users. There are risks to open architecture, as proven by the Cambridge Analytica scandals of 2016(Design et al., 2019).

7.2. Basic Principles of safety management

7.2.1. It is important to establish rules and guidelines for AI and user interactions. At the basis of these should be, a do no harm philosophy but also, a mandate to meet fundamental needs. This can be a complicated proposition as outcomes of AI analytics are often unpredictable.

7.2.2. Common link codes should contain, or be filtered, using a reference database of rules based on human law, but is able to adapt to unpredictable complexity.

7.2.3. Pathways should be kept open, and guidance used for most interactions, as blocking information that is potentially harmful could restrict development. It is better to perform risk analysis and offer advisory support, while ensuring that accountability is shared.

7.2.4. At the heart of safety principles lies risk analysis and mitigation. Human AI partnerships hold immense potential to expedite the examination of all hazard factors comprehensively, thereby eliminating most risks swiftly.

7.2.5. The abundance of collective data available enables an assurance of guideline adherence, efficient process implementation, and seamless integration of AI development with safety measures.

7.2.6. Risk analysis and mitigation is at the core of safety Principles. Human AI partnerships have the potential to eliminate most risk through thorough examination of all hazard factors quickly. There is an abundance of collective data that can be obtained to ensure guidelines can be followed, processes implemented, and AI development incorporates this.

7.3. Trust framework (TF)

7.3.1. Trust frameworks delineate the procedures, regulations, and technical requirements that safeguard the privacy, security, and integrity of data and transactions. They are instrumental in instilling confidence and certainty in the relationships and interactions among different entities engaged in online services and transactions.

7.3.2. In MNN Architecture a TF works as a means of authentication but also an intermediatory filter for specific needs only data transfer. A TF can be programmed to deliver based on previous exchanges with access modifications. An LLM enquiry can facilitate modifications based on information provided, compared against set rules without the need for further training guidance or input. Quick easy migration within shared sessions builds trust for human an AI alike.

8. Task displacement

8.1.1. A MNN augments human function by working in synergy with humans rather than displacing them entirely. While it's true that certain tasks can be automated and replaced by AI, the overall impact is not about eliminating jobs but optimizing productivity and resource allocation.

8.1.2. By automating repetitive and mundane tasks, MNN free up human resources to focus on more complex and creative endeavours. Humans are uniquely skilled in areas such as critical thinking, empathy, problem-solving, and emotional intelligence, which are not easily replicated by machines. Skills and experience gained still add value to ongoing tasks and can be re-utilised into advanced problem solving in crisis.

8.1.3. The increased productivity achieved through AI assistance allows humans to tackle more challenging problems, explore innovative solutions, and engage in strategic decision-making. It leads to a higher level of efficiency and effectiveness in various domains.

8.1.4. The time and resources saved by automating certain tasks can be reinvested in other areas that require human expertise and attention. This can lead to new job opportunities and skill development in fields that leverage uniquely human capabilities.

8.1.5. Rather than replacing humans, MNN augments human capabilities, enhancing their capacity to achieve more meaningful and impactful outcomes. It opens new possibilities for collaboration, where humans and AI work together to create a more prosperous and advanced society. The key lies in leveraging AI's strengths to complement human strengths, ultimately leading to a future where humans and machines coexist and thrive together.

9. Conclusion

(Conclusions are as per executive summary)

This paper was authored with the help of Chat GPT 3.5 and saved considerable hours through a human AI partnership. No humans were harmed during this process.

10. Citations used.

Chandler, D. (2013). Resilience and the Autotelic Subject: Toward a Critique of the Societalization of Security. *International Political Sociology*, *7*(2), 210–226. https://doi.org/10.1111/IPS.12018

Design, A. W.-S. J. T. J. of, Economics, undefined, and, undefined, & 2019, undefined. (n.d.). Design thinking for a user-centered approach to artificial intelligence. *Elsevier*. Retrieved August 4, 2023, from https://www.sciencedirect.com/science/article/pii/S2405872619300887

Mcleod, S. (2018). Maslow's Hierarchy of Needs. *Simply psychology*. https://www.simplypsychology.org/maslow.html

Mintz, S. (2015). *Do the Needs of the Many Outweigh the Needs of the Few? - Ethics Sage*. https://www.ethicssage.com/2015/03/do-the-needs-of-the-many-outweigh-the-needs-of-the-few.html

Montes, G. A., & Goertzel, B. (2019). Distributed, decentralized, and democratized artificial intelligence. In *Technological Forecasting and Social Change* (Vol. 141). https://doi.org/10.1016/j.techfore.2018.11.010

Moore, M., & Tambini, D. (2018). *Digital Dominance The Power of Google, Amazon, Facebook, and Apple*. https://lccn.loc.gov/2017052008