AI Horizons - Shaping Ethical Technologies for Human Dignity

IEEE AuDIITA 2024 REPORT

22 November 2024|Melbourne, Australia

2024 AuDIITA WORKSHOP

Dignity, Inclusion, Identity, Trust & Agency

22 Nov, 2024 | Melbourne, Australia

About IEEE DIITA

Many areas of human activity in the 21st Century take place within cyberspace. Those excluded from cyberspace are thereby excluded from a key domain of human endeavor. Exclusion may arise from many causes, including affordability, availability, discrimination, and concern for safety. “Dignity, Inclusion, Identity, Trust, and Agency” (DIITA) Industry Connections Program considers causes of exclusion which can be addressed by advancing technology for humanity through standardization and related solutions.

More Information Here:

https://standards.ieee.org/industry-connections/diita/

Thanks to Our AuDIITA 2024 Sponsors and Partners

KNOWLEDGE PARTNERS, SPONSORS & HOST

Disclaimer: The opinions expressed below are those of the speakers and participants, and do not necessarily reflect the views of IEEE, IEEE SA, or the IEEE SA DIITA activity.

Executive Summary: AI Horizons - Shaping Ethical Technologies for Human Dignity **4**

- AuDIITA Workshop 2024, Melbourne 4
- Workshop Intent 4
- Key Insights 4
- Key Recommendations for Stakeholders 5
- Methodology and Outcomes 6

Introduction **7**

Session 1: AI Systems in Defence Applications **10**

- Framework for Human Decision-Making in Autonomous Systems 11
- Scenario Analysis: Autonomous Weapons in a War Zone 12
- Framework Highlights & Key Insights 13
- Key Takeaways and Next Steps 13

Session 2: Domestic Violence and AI **14**

- Presentation Summaries 14
- AI and Domestic Violence Prevention 14
- Governance and Safeguards in AI 15
- Diversity and Inclusion in AI Development 17
- Breakout Discussions 18
- Key Takeaways 19
- Actionable Recommendations 20

Session 3: AI Trigger Points and Tangible Outcomes **20**

- Key Themes and Topics 21
- Key Discussion Points 22
- Recurring Themes 22
- Key Takeaways and Action Points 23

Plenary Session 4: Synthesizing Insights and Mapping the Way Forward **24**

Closing Remarks and Next Steps **25**

Conclusion **26**

Acknowledgement **29**

Appendix **31**

Executive Summary: AI Horizons - Shaping Ethical Technologies for Human Dignity

AuDIITA Workshop 2024, Melbourne

The AuDIITA Workshop 2024, hosted by Swinburne University of Technology, marked a significant advance in the dialogue around ethical AI development. Through its innovative format combining expert presentations with facilitated discussions, the workshop brought together technologists, ethicists, policymakers, and community representatives to explore how AI can better serve human dignity while addressing critical societal challenges.

Workshop Intent

The symposium moved beyond conventional discussions of AI ethics by creating spaces for genuine dialogue between diverse stakeholders. Through examining AI applications in defense and family violence prevention, participants explored the practical challenges of implementing ethical frameworks in high-stakes contexts. This approach revealed new insights about the intersection of technological capability, ethical imperatives, and social responsibility.

Key Insights

1. Human-Centric AI Development
    - The critical importance of human accountability in autonomous systems, particularly in defense applications
    - Recognition that AI should augment rather than replace human judgment in sensitive contexts
    - The imperative to address "digital erasure" of marginalized communities through inclusive design
    - The need for cultural change alongside technological solutions
2. Safety and Ethics Integration
    - Safety by Design principles must be embedded from the earliest stages of development
    - Robust governance frameworks should balance innovation with human rights protection
    - The tension between rapid technological advancement and ethical safeguards requires careful management

- Sustainable and circular approaches to AI development can enhance ethical outcomes

3. Inclusive Design and Implementation
    - Diverse perspectives are crucial in preventing bias and ensuring comprehensive representation
    - AI systems must reflect and respect various epistemologies, including indigenous perspectives
    - Marginalized communities must be engaged in both design and governance processes
    - Bio-inclusive and holistic approaches can guide more ethical development

Key Recommendations for Stakeholders

1. For Technology Developers
    - Integrate Safety by Design principles throughout the development lifecycle
    - Implement transparent AI processes with comprehensive documentation
    - Ensure diverse and representative datasets in AI training
    - Allocate equal resources to understanding social implications and technical development
    - Adopt sustainable and circular design approaches
2. For Policymakers
    - Develop agile regulatory frameworks that address emerging AI-related harms
    - Create legal mechanisms for developer accountability
    - Invest in understanding AI's societal implications through funded research
    - Review and adapt existing laws for AI's unique challenges
    - Ensure global regulatory agility in addressing emerging harms
3. For Industry Leaders
    - Foster cross-sector partnerships for sharing safety-enhancing tools
    - Balance transparency requirements with business confidentiality
    - Include diverse stakeholders in decision-making processes
    - Implement robust impact assessment protocols
    - Prioritize sustainable and ethical business models

4. For Academic and Research Institutions
    - Develop frameworks bridging theoretical ethics with practical implementation
    - Research AI's impact on marginalized communities
    - Promote interdisciplinary collaboration in AI development
    - Advance understanding of AI's role in addressing systemic challenges
    - Create evidence-based approaches to ethical AI governance

Methodology and Outcomes

The workshop employed a unique participatory methodology emphasizing collective wisdom and diverse perspectives. Each session combined expert presentations with facilitated small group discussions, captured through multiple channels including dedicated note-takers and participant documentation. This approach generated rich insights about developing AI systems that genuinely uphold human dignity while addressing practical implementation challenges.

The recommendations emerging from the workshop emphasize the need for collaborative, inclusive approaches that prioritize human dignity while acknowledging AI's transformative potential. These insights provide a foundation for stakeholders to move forward with practical, ethical implementations that serve the broader public interest while promoting responsible innovation.

Introduction

The IEEE Standards Association proudly presented the AuDIITA Workshop 2024 on November 22, 2024, under the core principles of the DIITA program. Hosted by Swinburne University of Technology and titled "AI Horizons: Shaping Ethical Technologies for Human Dignity," this dynamic workshop brought together leading experts, technologists, ethicists, and policymakers to explore the critical intersection of artificial intelligence (AI) and human dignity. Local volunteers of the IEEE Society on Social Implications of Technology played a key organising role in the event.

The workshop, held in the AGSE Building's AGSE202 room, at Swinburne University was structured around four key sessions: AI Systems in Defence Applications, Family Violence and AI, AI Trigger Points and Tangible Outcomes, and a final synthesis session. Each session was designed to foster deep engagement with complex ethical challenges through a unique participatory methodology that emphasized collective wisdom and diverse perspectives.

The day's format intentionally moved beyond traditional conference presentations to create spaces for genuine dialogue and collaborative insight generation. Each session began with thought-provoking presentations from expert panelists, followed by facilitated small group discussions where participants could deeply explore the issues raised. These breakout conversations were captured through multiple channels: dedicated volunteer note-takers documented the emerging themes, while participants recorded their own insights and reflections on butcher's paper and Post-it notes placed on discussion tables.

A skilled facilitator guided the day’s flow, ensuring that insights from each session were brought back to the larger group for collective consideration and building a progressive synthesis of understanding. This approach allowed the workshop to examine critical areas including AI’s applications in defense, its role in family violence prevention and identification, emerging ethical trigger points, and broader societal impacts. The methodology particularly emphasized the importance of hearing diverse voices and perspectives, ensuring that discussions around ethical AI development were grounded in real-world implications and human experience.

Through this carefully structured yet organic process, the workshop created a rich tapestry of insights about developing AI systems that genuinely uphold human

dignity. The discussions revealed complex intersections between technical capability, ethical imperatives, and social responsibility, while generating practical recommendations for moving forward in these critical areas.

Detailed Recommendations and Stakeholder Analysis

The workshop generated comprehensive recommendations that emerged from the day's rich discussions and diverse perspectives. These recommendations reflect both immediate practical needs and longer-term aspirational goals for ethical AI development.

Recommendations for Technology Developers

The integration of Safety by Design principles throughout the AI development lifecycle emerged as a foundational requirement, particularly in addressing technology-facilitated abuse and potential AI system misuse. This approach emphasizes prevention rather than mitigation of harm.

Key Stakeholders:

- Software development teams and product managers
- Technology company leadership and ethics committees
- Standards Australia and similar standards bodies
- eSafety Commissioner and safety-focused organizations
- Industry bodies focused on ethical tech development

Implementation Requirements:

- Systematic changes to development processes
- Enhanced testing protocols with specific focus on harm prevention
- Regular ethical audits and impact assessments
- Integration of diverse datasets and perspectives
- Adoption of sustainable and circular design approaches

Recommendations for Policymakers

The development of agile regulatory frameworks must balance addressing emerging AI-related harms while fostering beneficial innovation. This was particularly evident in discussions about autonomous systems in defense and family violence prevention technologies.

Key Stakeholders:

- Federal and state government agencies
- Legal experts and human rights organizations
- Technical standards bodies
- Public safety and law enforcement agencies
- Civil society organizations

Implementation Priorities:

- Creation of adaptive regulatory frameworks
- Development of accountability mechanisms
- Investment in research on societal implications
- Integration of human rights protections
- Establishment of cross-jurisdictional cooperation

Recommendations for Industry Leaders

Cross-sector partnerships and transparent practices emerged as critical for addressing the complex challenge of balancing commercial interests with ethical imperatives. This includes particular attention to bias prevention and diverse representation in AI development.

Key Stakeholders:

- Industry associations and peak bodies
- Corporate leadership teams
- Ethics advisory boards
- Research and development departments
- Community stakeholder groups

Implementation Strategies:

- Establishment of collaborative industry models
- Development of shared best practices
- Creation of industry-wide ethical standards
- Implementation of transparent AI processes
- Integration of sustainable business models

Recommendations for Academic and Research Institutions

The focus on developing practical frameworks that bridge theoretical ethics with implementation addresses the clear gap between academic discourse and industry practice. This includes the development of evidence-based approaches to AI governance.

Key Stakeholders:

- Universities and research institutes
- Funding bodies and grant agencies
- Academic-industry partnership programs
- Ethics review boards
- Research communities

Implementation Approaches:

- Development of practical assessment tools
- Creation of ethical implementation frameworks
- Establishment of interdisciplinary research programs
- Focus on marginalized community impacts
- Integration of diverse epistemologies

Cross-Cutting Implementation Considerations

Success in implementing these recommendations requires attention to several common elements:

Resource Management:

- Consistent funding and resource commitment
- Clear allocation of responsibilities
- Sustainable resource planning

Stakeholder Engagement:

- Regular consultation and feedback mechanisms
- Inclusive decision-making processes
- Transparent communication channels

Measurement and Accountability:

- Development of clear success metrics
- Establishment of monitoring mechanisms
- Regular review and adjustment processes

Sustainability and Ethics:

- Integration of circular economy principles
- Consideration of long-term societal impacts
- Balance of innovation with ethical safeguards

Understanding the Day's Dialogue: Session Analysis

The recommendations and insights presented thus far emerged from rich, nuanced discussions across four distinct but interconnected streams. Each session created a unique space for exploring complex ethical challenges in AI development, from autonomous weapons systems to technology-facilitated abuse. The detailed analysis that follows reveals how high-level principles translate into practical concerns and real-world implications.

In examining AI's role in defense applications, participants grappled with fundamental questions about human agency and machine autonomy. The domestic violence session uncovered critical insights about technology's dual capacity to both harm and protect vulnerable populations. Discussions about AI trigger points revealed emerging challenges that current frameworks fail to address, while the final synthesis session wove these threads together into actionable insights.

We encourage readers to explore the detailed session analyses that follow, as they provide crucial context for understanding how our recommendations emerged from specific examples, challenges, and opportunities identified by participants. The devil truly is in the details – seemingly straightforward principles like "Safety by Design" take on new meaning when examined through the lens of specific use cases and real-world scenarios discussed during these sessions.

The analysis of each session captures not only the formal presentations but also the valuable insights that emerged from group discussions, participant feedback, and collaborative problem-solving exercises. These ground-level perspectives provide essential context for understanding why certain recommendations were prioritized and how they might be implemented effectively.

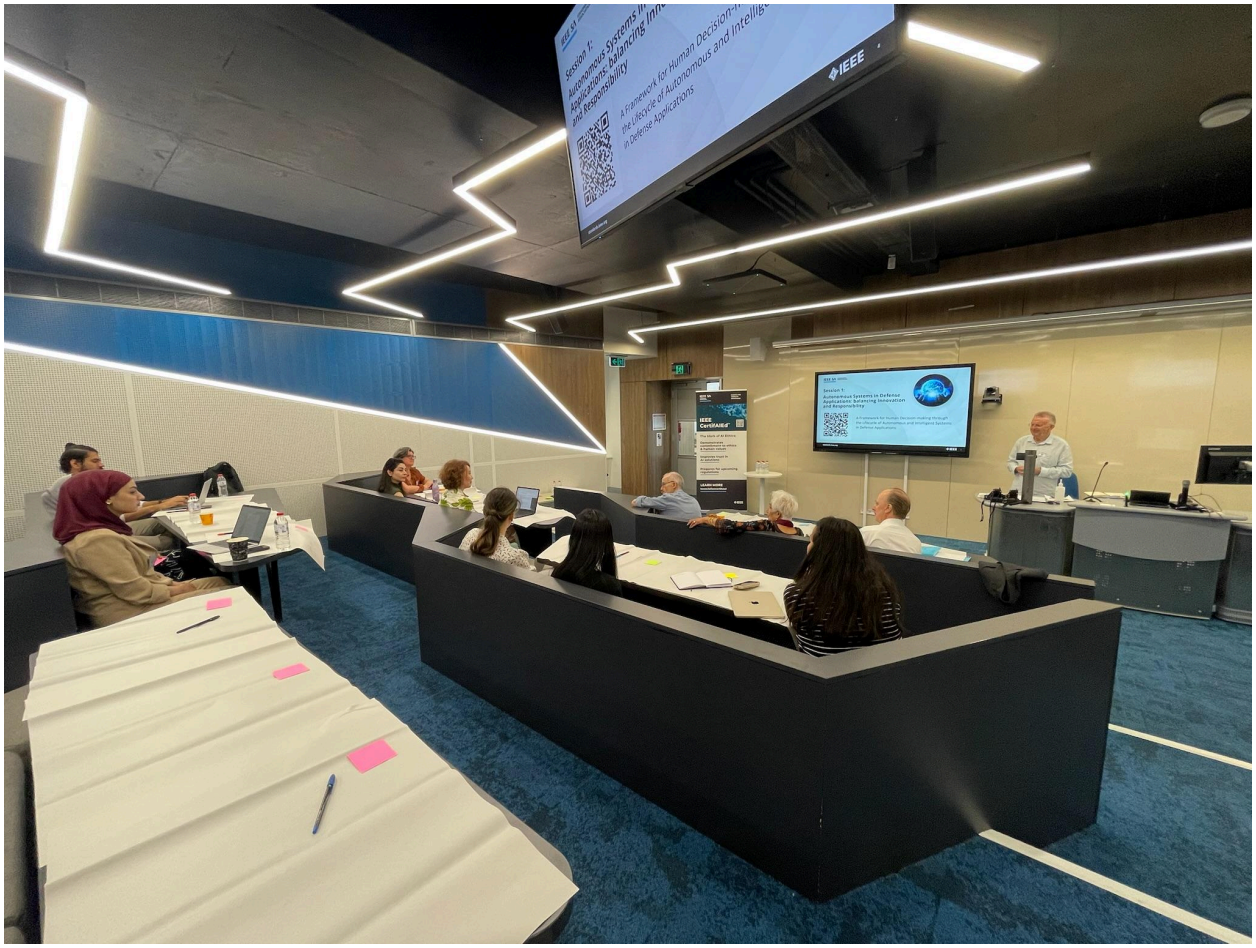
Session 1: AI Systems in Defence Applications

Led by: Dr. **Greg Adamson, Chief Information Security Officer, Department of Transport and Planning**

Background Summary

The first session focused on the **ethical considerations surrounding autonomous weapons and AI in defense**. Drawing from the recently released IEEE report A framework for human decision-making through the lifecycle of autonomous and intelligent systems in defense applications, the discussion highlighted the evolution of AI ethics frameworks, emphasizing the importance of human rights in technology design. The session

Reviewed Australia's AI ethics principles, particularly the initial two principles derived from IEEE's ten overarching guidelines and explored the transition from a utilitarian approach to more human-centric ethical considerations.



Discussion Highlights

Framework for Human Decision-Making in Autonomous Systems

Greg presented a comprehensive framework for integrating human decision-making within autonomous and intelligent systems in defense contexts. He traced the historical development of ethical AI frameworks, referencing the IEEE's contributions and Australia's adoption of these principles. The discussion underscored the necessity for human responsibility and accountability, particularly in scenarios where AI systems could potentially cause harm, such as autonomous weapons engaging in combat.

About the Framework

The framework set out in this document is intended to help inform a range of stakeholders involved in various types of decision-making related to autonomous and intelligent systems (AIS), especially in the military context. The framework provides a granular means for identifying the technical considerations that are necessary to meet legal requirements, to inform and guide ethical discussions,

and to determine human responsibility and accountability across the entire lifecycle of AIS.

Scenario Analysis: Autonomous Weapons in a War Zone

A thought-provoking scenario was introduced where an autonomous weapon system is deployed to protect a humanitarian ship carrying grain in a war zone. The scenario raised critical ethical questions:

- Should autonomous weapons be deployed in theaters of war?
- Can such systems ensure respect for human dignity and avoid unintended harm?

Military experts with participants provided legal insights, advocating for the prohibition of weaponizing general-purpose robots and ensuring that AI systems do not learn and adapt in the field without rigorous testing and evaluation. The consensus highlighted the importance of establishing standardized protocols for autonomous weapons, emphasizing interoperability and shared ethical standards among different manufacturers and nations.

Framework Highlights & Key Insights

- **Human Responsibility and Accountability:** Central to the ethical deployment of AI in defense, ensuring that humans remain accountable for AI-driven actions.
- **Inclusivity in Ethical Frameworks:** Incorporating diverse perspectives to create comprehensive and inclusive ethical guidelines.
- **Standardization:** Developing universally accepted standards for autonomous weapons to prevent misuse and ensure interoperability.
- **Risk Mitigation:** Addressing potential failures in AI operability and ensuring robust protective frameworks.

Key Takeaways and Next Steps

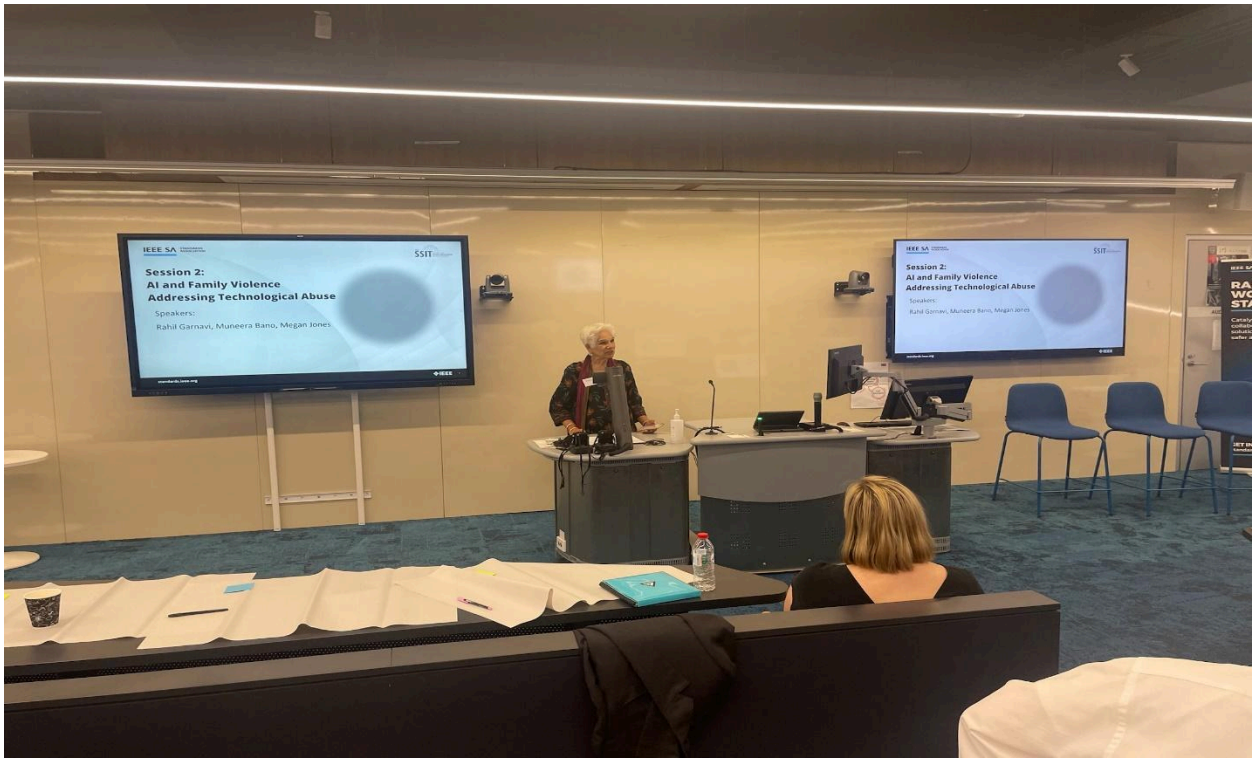
- **Ethical Boundaries:** Consensus that general-purpose robots, especially those intended for civilian use, should not be weaponized.
- **Human Rights Prioritization:** Ensuring that human rights remain at the core of any AI ethical framework.
- **Collaborative Efforts:** Emphasizing the need for multidisciplinary collaboration among technologists, ethicists, and legal experts to navigate the ethical complexities of AI in defense.

Session 2: Domestic with Family Violence and AI

Led by: Professor Supriya Singh, Adjunct Professor with the Department of Social Inquiry, La Trobe University

Presenters:

- Megan Jones, Industry Engagement, eSafety Commissioner, Leading eSafety's Safety by Design initiative
- Professor Rahil Garnavi, Director, RAIsE Hub, RMIT University
- Dr. Muneera Bano, Principal Research Scientist, Co-Lead AI Diversity and Inclusion, CSIRO Data61



Presentation Summaries

AI and Domestic Violence Prevention

Megan Jones highlighted the intersection of AI and domestic violence, focusing on how AI can both exacerbate and mitigate abuse. Key points included:

- **Technology-Facilitated Abuse:** AI enables but can also be embedded and implicated in the rapid dissemination of harmful content like doxing, threats, and deepfake pornography, predominantly targeting women.
- **Safety by Design:** Emphasizing the integration of safety measures during AI development to minimize harms, including service provider responsibility and user empowerment.
- **Proactive Solutions:** Tools like StopNCII.org that detect and remove non-consensual intimate images using privacy-preserving technologies.
- **Accountability and Collaboration:** The necessity for industry collaboration to create safer digital environments that prioritize user safety and dignity.

Governance and Safeguards in AI

Rahil Garnavi discussed the importance of robust governance frameworks to prevent AI misuse in family violence contexts. Key insights included:

- **Case Studies:** Examination of Spain's VioGén system, highlighting the risks of over-reliance on AI in violence detection. See reported analysis presented below for further insights on this scenario.
- **Exploitation of Technologies:** Smart home and social media technologies can be misused by perpetrators for harassment and control.
- **Design Principles:** Advocating for AI systems that identify abuse patterns without reinforcing biases, requiring human oversight and continuous monitoring.
- **Governance:** The need for transparent AI processes and ongoing monitoring to ensure fairness and inclusivity.

Case study analysis - VioGen

According to research by AlgorithmWatch (2020), Spain's VioGén system represents both the potential and challenges of using AI in domestic violence prevention. As reported by Michele Catanzaro, VioGén is considered the world's most comprehensive algorithmic system for assessing gender violence risks, having processed over 600,000 cases by January 2020.

The system works by having police officers complete a risk assessment form with 39 items, covering factors such as previous assault severity, aggressor characteristics, victim vulnerability, and aggravating circumstances. This data feeds into an algorithm that calculates a risk score, which then determines the level of protection measures implemented.

While the system demonstrates reasonable effectiveness (with performance metrics comparable to medical screening tests), several critical challenges have emerged:

1. **Implementation Gaps**: Despite the 2004 law requiring interdisciplinary teams including psychologists and social workers, resource constraints have limited this comprehensive approach.
2. **Training Deficiencies**: Many police officers lack specialized training in gender-based violence, potentially affecting the quality of data input and risk assessment interpretation.
3. **Risk Assessment Limitations**: Critics, particularly victims' organizations, argue that categorizing any case as "low risk" may be fundamentally flawed, as the act of reporting itself often escalates risk.
4. **System Evolution**: Following tragic failures, including cases where "low-risk" assessments preceded homicides, the system has undergone multiple iterations. The latest version (VioGén 5.0) includes specific protocols for cases involving children and potential lethal outcomes.

This case study illustrates the complex challenges of implementing AI systems in highly sensitive contexts where human lives are at stake, highlighting the critical importance of continuous system refinement, comprehensive training, and careful consideration of how algorithmic assessments interact with human judgment.

Diversity and Inclusion in AI Development

Muneera Bano, emphasized the critical role of diversity and inclusion in AI to prevent with"digital erasure" with marginalized groups. Key points included:

- **Personal Experiences:** Shared personal stories illustrating the exclusion of women in data and digital systems.
- **Promoting Diversity:** Highlighted CSIRO's efforts to incorporate diversity in AI across various sectors to address biases.
- **Inclusive Design:** Ensuring AI systems represent diverse communities to meet the needs of all users.
- **Continuous Governance:** Advocated for ongoing inclusive design and governance processes to align AI systems with societal contexts and human rights.

Breakout Discussions

Participants engaged in breakout group discussions focusing on themes such as transparency, representation, ethical versus business priorities, and contextual considerations. Notable insights from group discussions included:

- **Transparency vs. Business Confidentiality:** Balancing openness with the need to protect business interests.
- **Representation and Inclusivity:** Ensuring AI systems adequately represent and support marginalized communities.
- **Economic vs. Ethical Priorities:** Addressing the tension between rapid technological advancement and ethical safeguards.
- **Practical Applications:** Exploring how AI can provide immediate support to victims without replacing essential human intervention.

Key Takeaways

- **Potential and Risks of AI:** AI can support victims through discreet tools like chatbots but poses risks of misclassification and reinforcement of harmful stereotypes.
- **Inclusive Data Training:** AI must be trained on diverse datasets to reflect the experiences of all communities, preventing "digital erasure."
- **Robust Governance Frameworks:** Essential for ensuring AI systems are fair, transparent, and accountable.
- **Safety Measures:** Implementing safeguards like watermarking synthetic content to reduce misuse and build trust.
- **Collaboration:** Essential partnerships between industry professionals, policymakers, and affected communities to develop ethical AI solutions.

Actionable Recommendations

- **Embed Safety Measures:** Integrate Safety by Design principles during AI development to anticipate and mitigate risks.
- **Prioritize Inclusivity:** Engage marginalized groups in the AI design process to address their unique needs.
- **Continuous Assessment:** Regularly evaluate AI systems for fairness, inclusivity, and unintended consequences.
- **Update Datasets:** Ensure datasets are diverse and representative of all communities.
- **Foster Partnerships:** Collaborate across sectors to share safety-enhancing tools and best practices.
- **Advocate for Supportive Policies:** Push for funding and policies that prioritize marginalized communities in AI development.
- **Document AI Processes:** Promote initiatives like IBM's Fact Sheets to maintain transparency in AI development.
- **Ensure Legal Accountability:** Develop legal frameworks to hold developers accountable for AI misuse.

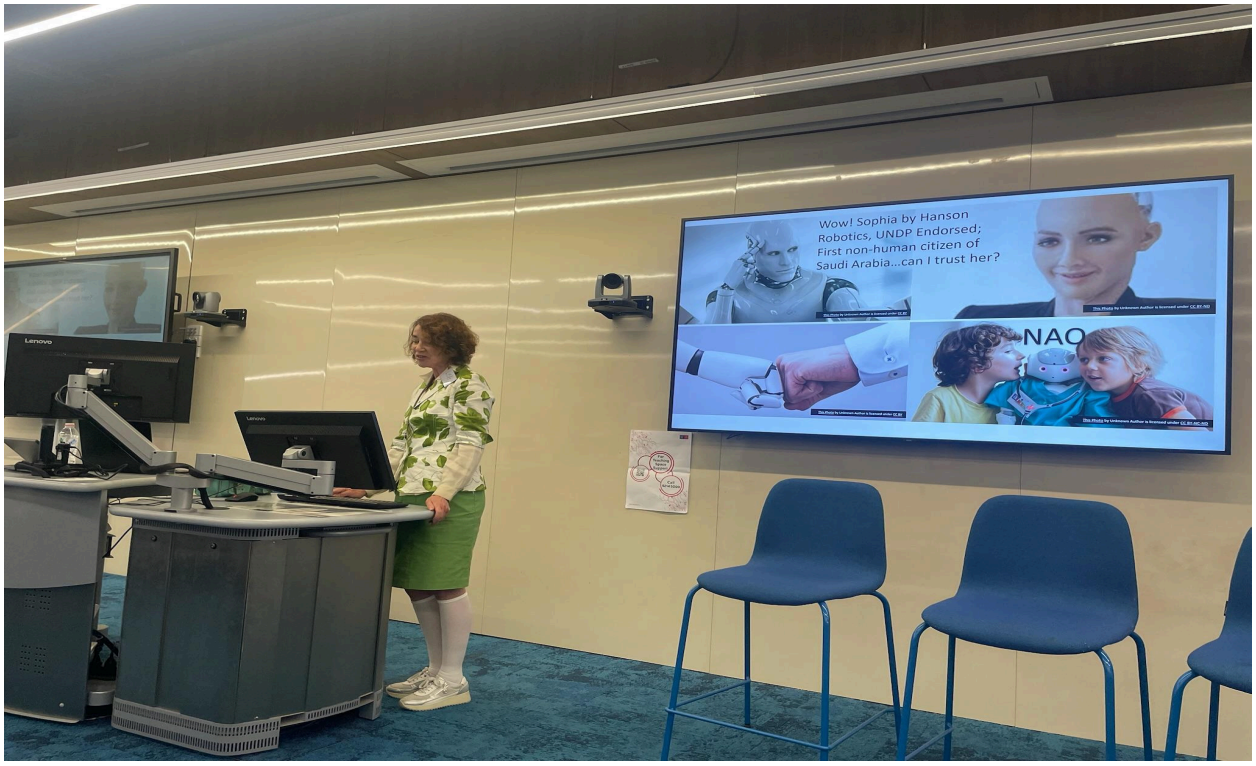
Session 3: AI Trigger Points and Tangible Outcomes

Facilitators

- **Dr Greg Adamson, Chief Information Security Officer, Department of Transport and Planning**
- **Dr Alexia Maddox, Director of Digital Education, School of Education, La Trobe University**

Presenter & Speaker: Dr Lisa Dethridge, RMIT University

Overview:
This session explored the interplay between technological innovation and ethical safeguards, emphasizing the need for governance frameworks, inclusivity, and responsible AI standards. The discussions revolved around practical recommendations for addressing trigger points in AI ethics.



Key Themes and Topics:

1. **Standards for Responsible AI Use and Governance:**
    - The session highlighted the importance of standards and governance frameworks that balance innovation with ethical imperatives.

2. **Balancing Innovation with Ethical Safeguards:**
    - Stephanie and Megan discussed how regulatory frameworks should focus on mitigating worst-case harms while allowing standards to encourage long-term, innovative development.
    - Megan stressed that companies should dedicate more time to testing the societal impact of technologies pre-launch.

3. **Inclusivity and Broader Ethical Considerations:**
    - Discussions emphasized bio-inclusivity and a circular/holistic approach to technology.
    - The need for diverse voices, including marginalized communities, in shaping AI frameworks was underscored, as raised by Parisa and Fayr.

Key Discussion Points:

- Megan shared insights from her experience as a government regulator, emphasizing that tech companies often push for clear regulatory frameworks to avoid stakeholder-driven dilemmas.
- Natasha highlighted India's TikTok ban, raising concerns about AI's societal impact and regulatory agility.
- David A. proposed the nationalization of companies acting in bad faith over time, suggesting institutional change driven by robust standards.

Recurring Themes:

- The tension between economic interests and ethical considerations, with Lisa remarking that unregulated human valuation of capital keeps society stuck in antiquated models.
- The need to center human concerns over machine efficiencies, with Greg cautioning against overestimating AI's capabilities and underestimating human ingenuity.
- Calls for leveraging AI's strengths in data analysis (e.g., court systems and medical diagnostics) while addressing ethical challenges like surveillance and bias.

Key Takeaways and Action Points:

- Align AI with sustainability goals, focusing on systemic issues like climate change and poverty alleviation.
- Develop clear legal and ethical frameworks to guide AI deployment and prevent misuse.
- Prioritize inclusivity and avoid digital exclusion by addressing data biases and designing tools for marginalized groups.
- Advocate for transparency and oversight mechanisms to balance innovation with public interest.

Plenary Session 4: Synthesizing Insights and Mapping the Way Forward

Facilitator: Dr. Alexia Maddox, Director of Digital Education, School of Education, La Trobe University

Panel participants

- **Professor Supriya Singh, Adjunct Professor with the Department of Social Inquiry, La Trobe University**
- **Megan Jones, Industry Engagement, eSafety Commissioner, Leading eSafety's Safety by Design initiative**
- **Professor Rahil Garnavi, Director, RAIsE Hub, RMIT University**
- **Dr Stéphanie Camaréna, Founder and Principal - Source Transitions, Victoria Chapter Coordinator - SSIT**
- **Dr Greg Adamson, Chief Information Security Officer, Department of Transport and Planning**

Overview:
This plenary session synthesized the day's discussions, emphasizing the critical role of human-centric design, inclusivity, and actionable frameworks for responsible AI governance.

Highlights:

1. **Human-Centric Design:**
    - Lisa and Greg emphasized the necessity of focusing resources on educating humans about AI's risks and limitations rather than just training AI systems.
    - Supriya pointed out AI's potential to address gaps in systems, such as supporting survivors of financial abuse in under-regulated environments.
2. **Solutions Beyond Western Binaries:**
    - Panelists discussed incorporating diverse epistemologies, including indigenous perspectives, to break free from binary views of technology as inherently good or bad.

    - Greg reminded participants of technology's long-standing role in human evolution, advocating for a balanced and historical perspective on its impacts.
3. **Education and Regulation:**
    - Calls were made for stronger regulatory frameworks and greater investment in understanding AI's societal implications.
    - Megan emphasized the need for larger regulatory tools to hold companies accountable, while Stephanie suggested allocating equal resources to AI's social consequences as to its product development.

Action Planning Suggestions:

- Lisa proposed industrial design requirements to bake sustainability, inclusion, and diversity into every development stage.
- Supriya highlighted AI's potential to evidence financial abuse, advocating for careful implementation to avoid unintended consequences.
- Stephanie recommended involving stakeholders from diverse backgrounds in corporate conversations to ensure ethical AI implementation.

Closing Remarks and Next Steps

Recommendations

- Focusing on human-centric solutions and ethical AI integration.
- Revisiting pre-existing laws to adapt them for AI's unique challenges.
- Ensuring global regulatory agility to address emerging AI-related harms effectively.

Reflections and Final Thoughts

AI is a transformative tool with immense potential to address global challenges like inequality and climate change. However, it must be developed responsibly, with inclusivity, transparency, and accountability at its core. Embedding ethical considerations in AI's design and governance will ensure it empowers humanity rather than exacerbates societal disparities.

Conclusion

The AuDIITA Workshop 2024, “AI Horizons: Shaping Ethical Technologies for Human Dignity”, marked a significant step forward in our collective approach to ethical AI development. Through its innovative dialogue-driven format, the workshop demonstrated that meaningful progress in AI ethics requires more than just technical solutions or policy frameworks – it demands the integration of diverse perspectives, lived experiences, and collaborative wisdom.

The day’s discussions revealed several critical insights that challenge conventional approaches to AI development. First, the complex interplay between AI applications in defense and domestic violence prevention highlighted how ethical considerations must be contextual and nuanced, rather than universal and abstract. The workshop's examination of these areas demonstrated that effective AI governance requires careful attention to both immediate practical impacts and longer-term societal implications.

Particularly noteworthy was the emergence of a more nuanced understanding of human dignity in relation to AI systems. Participants moved beyond simplistic notions of AI safety to explore deeper questions about representation, inclusion,

and agency. The discussions around “digital erasure” and technology-facilitated abuse highlighted how AI systems can either perpetuate or help dismantle existing social inequities, depending on their design and implementation.

The workshop’s methodology proved particularly valuable in generating practical, actionable insights. By creating space for dialogue between technologists, ethicists, policymakers, and community representatives, the discussions revealed new possibilities for addressing complex challenges. The emphasis on capturing diverse perspectives through multiple channels – from expert panels to participant-driven discussions – enabled the emergence of rich, multifaceted solutions that bridge theoretical frameworks with practical implementation.

Looking forward, several key imperatives emerge from the day's proceedings:

- The need to embed ethical considerations and Safety by Design principles from the earliest stages of AI development
- The importance of creating governance frameworks that are both robust and adaptable
- The critical role of diverse voices and experiences in shaping AI systems that truly serve all of humanity
- The necessity of balancing innovation with careful consideration of societal impacts

Perhaps most significantly, the workshop demonstrated that meaningful progress in ethical AI development requires ongoing dialogue and collaboration across sectors and disciplines. The insights generated through this collaborative approach suggest that the path forward lies not in choosing between innovation and ethics, but in finding ways to advance both simultaneously through careful, inclusive, and thoughtful development practices.

As we move forward in the rapidly evolving landscape of AI technology, the principles and insights generated during this workshop provide a valuable foundation for future development. They remind us that the goal is not simply to create more advanced AI systems, but to ensure these systems enhance human dignity and contribute positively to society's well-being. The rich dialogue and practical recommendations that emerged from the workshop offer a promising pathway toward this goal, while acknowledging the ongoing work required to realize it fully.

From Dialogue to Action

The insights and recommendations emerging from this workshop represent not just aspirational goals, but a call to meaningful action. While this report captures the rich dialogue and critical insights from the day, the real work begins with translating these recommendations into concrete initiatives. We encourage stakeholders across technology development, policy making, industry leadership, and academic research to use these findings as a foundation for developing specific action plans within their spheres of influence. The IEEE DIITA program stands ready to support these efforts through continued dialogue, resource sharing, and community building. Those interested in taking forward any of the recommendations outlined in this report are encouraged to connect with the IEEE DIITA community to explore collaborative opportunities for implementation.

Contact: Dignity, Inclusion, Identity, Trust, and Agency activity interest form

Acknowledgements

The success of the AuDIITA Workshop 2024 was made possible through the dedicated efforts of many individuals who contributed their time, expertise, and insights throughout the day, along with our sponsors.

We extend our sincere gratitude to our volunteer note-takers who diligently captured the rich discussions and emerging insights:

- Dr Romaine Logere
- Nishanth Nagarajan
- Parisa Sekandari, Multicultural Women's Alliance Against Family Violence
- Russell Jana Mariano, Multicultural Women's Alliance Against Family Violence

- Devaka Basnet, Multicultural Women's Alliance Against Family Violence

Their careful attention to detail and commitment to documenting the day's proceedings has been invaluable in capturing the workshop's valuable conversations and outcomes.

We would also like to express our deep appreciation to our distinguished speakers, panelists and session facilitators who brought their expertise and thought leadership to the workshop:

- Professor Supriya Singh, Adjunct Professor with the Department of Social Inquiry, La Trobe University
- Megan Jones, Industry Engagement, eSafety Commissioner, Leading eSafety's Safety by Design initiative
- Professor Rahil Garnavi, Director, RAIsE Hub, RMIT University
- Dr Muneera Bano, Principal Research Scientist, Co-Lead AI Diversity and Inclusion, CSIRO Data61
- Dr Stéphanie Camaréna, Founder and Principal - Source Transitions, Victoria Chapter Coordinator - SSIT
- Dr Greg Adamson, Chief Information Security Officer, Department of Transport and Planning
- Dr Alexia Maddox, Director of Digital Education, School of Education, La Trobe University

Their provocative insights and weighty considerations created fertile ground for meaningful dialogue and helped participants engage deeply with complex ethical challenges in AI development. Their contributions have significantly advanced our collective understanding of how to shape ethical technologies that truly uphold human dignity.

The thoughtful engagement of all participants, combined with the expertise of our speakers and the dedication of our note-takers, has created a valuable resource for continuing discussions about ethical AI development and implementation.

Our sponsors

We extend our sincere gratitude to all organizations and partners who contributed to the success of the AuDIITA workshop. Their support and collaboration were instrumental in bringing this important event to fruition.

We particularly acknowledge IEEE for their comprehensive support as a primary host, including administrative assistance, promotion, and reporting that were vital to the workshop's success.

Special thanks to Swinburne University of Technology for hosting the event and providing both the venue and essential support services throughout the workshop.

We are deeply grateful to our knowledge partners whose support, expertise and networks enriched the workshop's contribution and reach:

- Multicultural Women's Alliance Against Family Violence
- IEEE Victorian Section
- Society on Social Implications of Technology (IEEE) for their valuable contribution to workshop ideation and speaker identification
- IEEE Digital Privacy
- Internet Society
- La Trobe University

Their collective contributions and commitment to this initiative have been invaluable in advancing our workshop's objectives and fostering meaningful discussions in this important field.

Appendix

Key Resources and Reference Materials

This appendix compiles essential resources referenced during the workshop and materials that support the implementation of key recommendations.

Frameworks and Standards

- IEEE Ethically Aligned Design Framework
    - Core principles for ethical AI development
    - Guidelines for implementation Link: standards.ieee.org/industry-connections/diita/
- Safety by Design Initiative (eSafety Commissioner)
    - Practical guidelines for technology development
    - Assessment tools and implementation frameworks Link: esafety.gov.au/industry/safety-by-design

Case Studies and Implementation Examples

- VioGén System (Spain)
    - Comprehensive analysis by AlgorithmWatch (2020)
    - Lessons learned and system evolution Link: Automating Society
- StopNCII.org
    - Technology-facilitated abuse prevention
    - Privacy-preserving detection tools Link: stopncii.org

Professional Development and Training Resources (eSafety)

eSafety Commission Educational Resources

The eSafety Commission provides comprehensive training and educational resources for various stakeholders working in family violence prevention and response:

For Educators and Professional Staff

- Professional learning program for educators and school staff
- Curriculum-aligned resources and classroom guidance

- Professional standards-aligned training modules Visit: Training for educators | eSafety Commissioner

Frontline Workers Training Portal

- Specialized online training for those working directly with vulnerable populations
- Modules on technology-facilitated abuse recognition and response
- Resources for supporting clients experiencing digital safety issues Access: online training for frontline workers

Community Organization Resources

- Training materials for community groups and organizations
- Workshops and presentation materials
- Resources for community education and awareness Learn more: Community education and training | eSafety Commissioner

These resources provide practical tools and knowledge for implementing safety-first approaches when working with AI and digital technologies in contexts where family violence is a concern. The training materials are regularly updated to reflect emerging technologies and threats.

Research and Publications

- S. Allik; R. Azafrani; D. Barnes; I. Bode; A. Conn; E. Lach M; C. Lennon; R. Liivoja; J. Shanahan; M. Stephens; L. Titus; A. Wagner R, "A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications," in *A Framework for Human Decision-Making through the Lifecycle of Autonomous and Intelligent Systems in Defense Applications* , vol., no., pp.1-63, 4 Oct. 2024. https://ieeexplore.ieee.org/document/10707139

Further Reading

Selected publications recommended by workshop presenters addressing key themes of human dignity, inclusion, and ethical AI development.

AI and Family Violence

Crooks, C. L., Talwalkar, S., Sharma, T., Arora, K., & Venkatesh, K. (2024, 15-24 March 2024). Designing Human-centered Artificial Intelligence to Assist with Domestic Abuse Recovery: Mitigating Technology Enabled Coercive Control. SoutheastCon 2024,

Garcia-Vergara, E., Almeda, N., Fernández-Navarro, F., & Becerra-Alonso, D. (2023). Artificial intelligence extracts key insights from legal documents to predict intimate partner femicide. Scientific reports., 13(1), 18212.

Novitzky, P., Janssen, J., & Kokkeler, B. (2023). A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools. Heliyon, 9(6), e17140.

Petering, R., Um, M. Y., Fard, N. A., Tavabi, N., Kumari, R., & Gilani, S. N. (2018). Artificial intelligence to predict intimate partner violence perpetration. Artificial intelligence and social work, 195.

AI and Ethics

Seyfert, R., & Roberge, J. (2016). Algorithmic cultures: Essays on meaning, performance and new technologies. Taylor & Francis.

Seyfert, R. (2021). Algorithms as regulatory objects. *Information, Communication & Society*, *25*(11), 1542–1558. https://doi.org/10.1080/1369118X.2021.1874035

Rathenau Institute Manifesto (2020) on digital rights and public spaces: Rathenau Manifesto

Mathews, F. (2011). Towards a Deeper Philosophy of Biomimicry. Organization & Environment, 24(4), 364-387. https://doi.org/10.1177/1086026611425689

Australian Government frameworks

- Australian Government et al. (2024) National framework for the assurance of artificial intelligence in government, Australian Government, accessed 5 December 2024 National framework for the assurance of artificial intelligence in government
- Australia's AI Ethics Principles
- Department of Home Affairs Security of Critical Infrastructure (SoCI)
- ACMA regulations
- National Health and Medical Research Council Codes of Ethics

For additional resources or to contribute to this knowledge base, please contact Dignity, Inclusion, Identity, Trust, and Agency activity interest form

Report : Compiled and developed by

Nitin Kumar

Program Manager

IEEE SA, India Office