The Evolution of Cybersecurity Research at CSIRO: A Two-Decade Journey and Future Outlook

Dr. Surya Nepal
Group Leader, Cybersecurity and Quantum Systems
CSIRO's Data61

Australia's National Science Agency

I would like to begin by acknowledging the Traditional Owners of the land on which we’re meeting today and pay my respects to their Elders, past and present.

I would like to acknowledge and thank all PhD students and the research and engineering staff who contributed to this work. All work presented here was done at CSIRO, where I am involved in some capacity.

CSIRO’s focus areas

- Agriculture and Food
- Energy
- Health and Biosecurity
- Environment
- Manufacturing
- Mineral Resources

Data61

- Space & Astronomy
- Australian Centre for Disease Preparedness (ACDP)
- Marine Facility
- National Computing Infrastructure
- Research Collections

Big ideas start here

Fast WIFI

PLASTIC BANKNOTES

AEROGARD

BARLEYmax™

RELENZA FLU TREATMENT

TOTAL WELLBEING DIET

HENDRA VACCINE

EXTENDED WEAR CONTACTS

SOFTLY WASHING LIQUID

SELF TWISTING YARN

RAFT POLYMERISATION

NOVACQ™ PRAWN FEED

CSIRO's Data61: Australia's Largest Data & Digital Innovation R&D Organisation

1000+
talented people
(including affiliates/students)

300+
PhD students
30+
University collaborators

200+
Gov & Corporate partners

Data61 Generated
18+ Spin-outs
130+ Patent groups

AI

Responsible AI

Privacy & RegTech

Engineering & Design of AI Systems

Resilient & Recovery Tech

Cybersecurity

Digital Twin

Spark (bushfire) toolkit

Facilities

Mixed-Reality Lab

Robotics Inno. Centre

AI4Cyber HPC Enclave

Research Capabilities in Data61

Cyber Physical Systems

- Autonomous robotics
- AI enabled computer vision
- 3D mapping
- Distributed sensing

Analytics & Decision Sciences

- Machine learning
- Quantitative risk assessment
- Computational linguistics
- Market design

Software & Computational Systems

- **Security, Privacy, Critical infra.**
- **(Responsible) AI Engineering**
- **Computational, Data and Analytics Platforms**
- **Quantum systems/security**

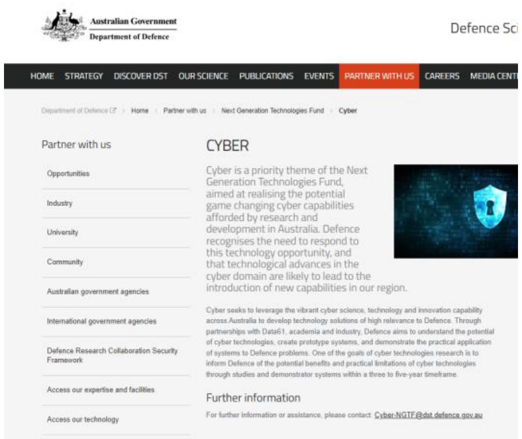
Cybersecurity and Quantum System Group

Capabilities

- **60+** Scientists and **50+** PhD students.
- **4** Teams at the intersection of AI, Cybersecurity, Human-centric and Quantum.
- Ranked in the **top 10** worldwide in terms of Scientific publications at the top Cybersecurity Conferences.
- **20+ externally** funded projects with national and international partners.

Strategic Partnership

- **AU**: DSTG group, US Army, AU Army, ASCA, DHA, Cybersecurity CRC, Gov agencies, AU universities.
- **USA**: Purdue, Indiana, Georgia Tech, Uni of Pittsburgh. DHS
- **UK**: Alan Turing, Newcastle, Cardiff.
- **Singapore**: SMU, NTU, A*STAR.
- **Korea**: ETRI, SKKU.
- **Industries**: NVIDIA, Google, Penten, xAmplify, etc.



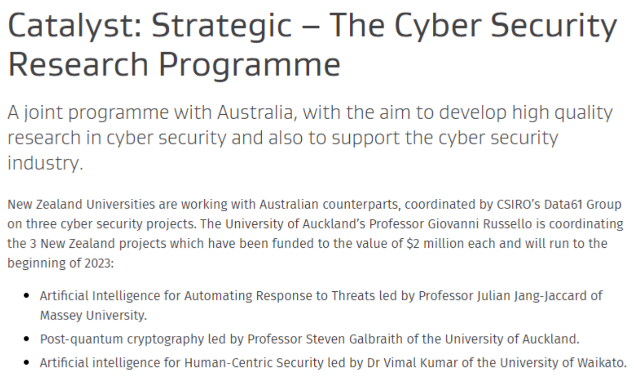
CYBER

Cyber is a priority theme of the Next Generation Technologies Fund, aimed at realising the potential game changing cyber capabilities afforded by research and development in Australia. Defence recognises the need to respond to this technology opportunity, and that technological advances in the cyber domain are likely to lead to the introduction of new capabilities in our region.

Cyber seeks to leverage the vibrant cyber science, technology and innovation capability across Australia to develop technology solutions of high relevance to Defence. Through partnerships with Data61, academia and industry, Defence aims to understand the potential of cyber technologies, create prototype systems, and demonstrate the practical application of systems to Defence problems. One of the goals of cyber technologies research is to inform Defence of the potential benefits and practical limitations of cyber technologies through studies and demonstrator systems within a three to five-year timeframe.

Further information

For further information or assistance, please contact: Cyber-NGTF@dst.defence.gov.au

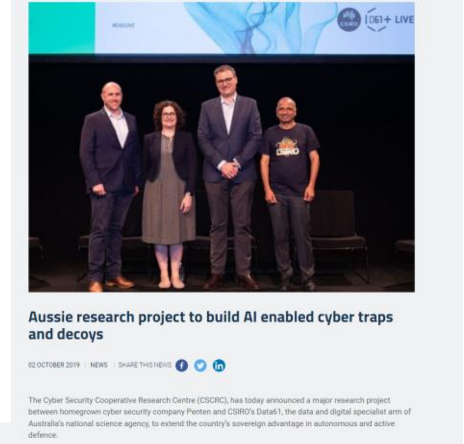


Catalyst: Strategic – The Cyber Security Research Programme

A joint programme with Australia, with the aim to develop high quality research in cyber security and also to support the cyber security industry.

New Zealand Universities are working with Australian counterparts, coordinated by CSIRO's Data61 Group on three cyber security projects. The University of Auckland's Professor Giovanni Russello is coordinating the 3 New Zealand projects which have been funded to the value of $2 million each and will run to the beginning of 2023:

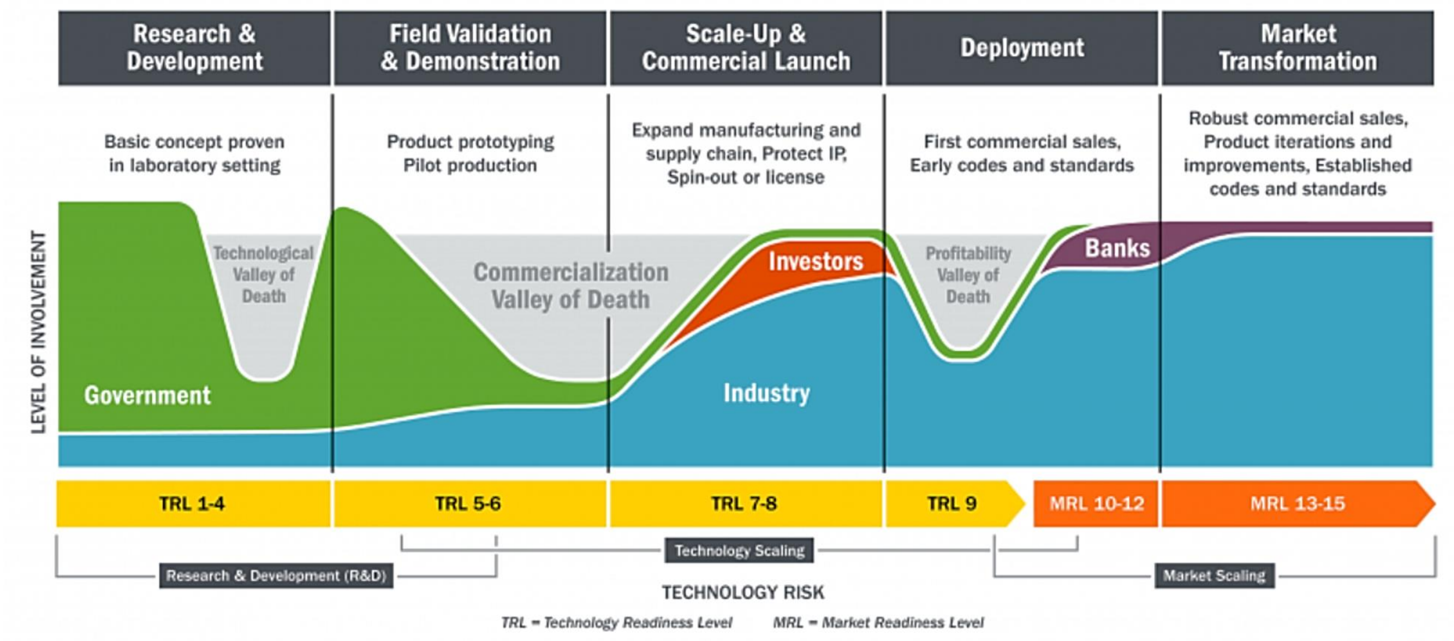
- Artificial Intelligence for Automating Response to Threats led by Professor Julian Jang-Jaccard of Massey University.
- Post-quantum cryptography led by Professor Steven Galbraith of the University of Auckland.
- Artificial intelligence for Human-Centric Security led by Dr Vimal Kumar of the University of Waikato.



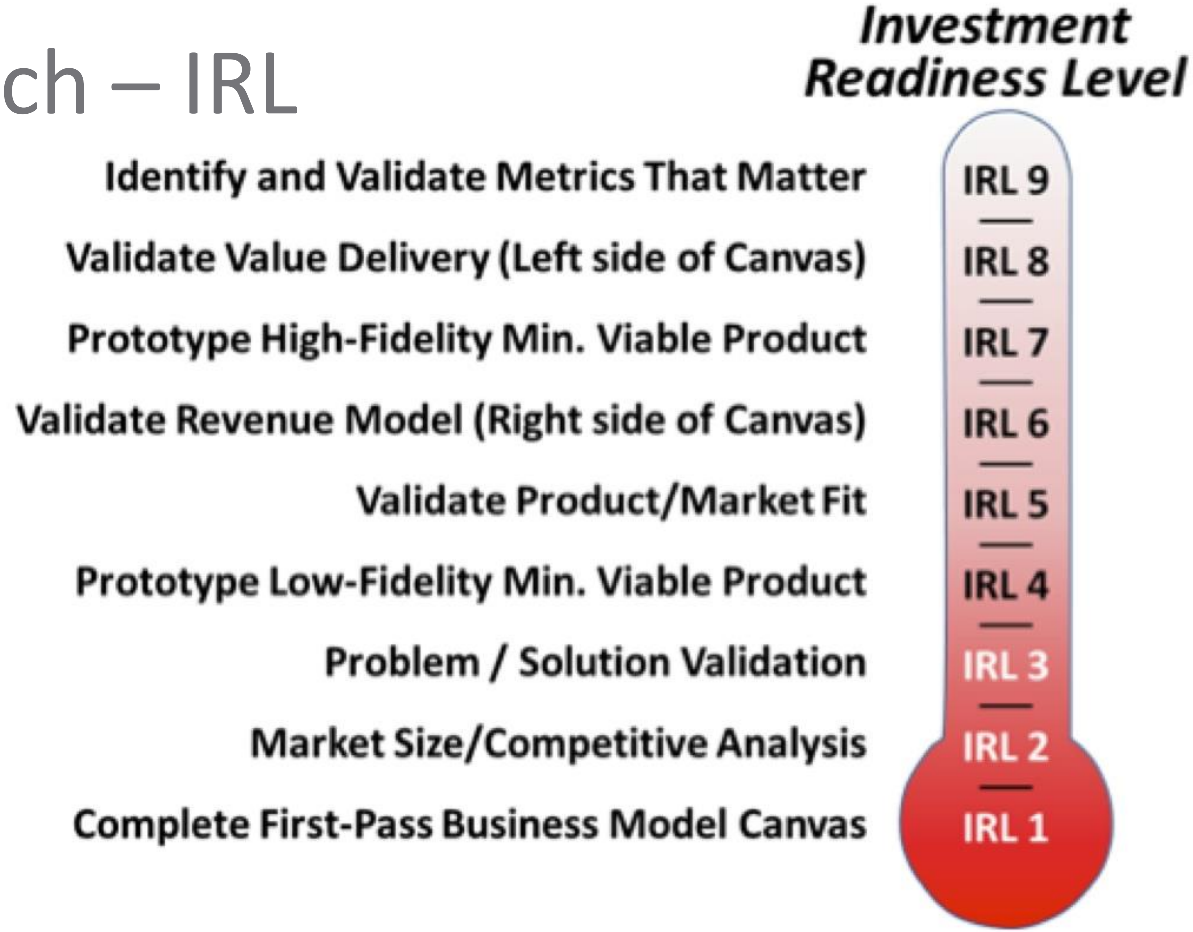
Aussie research project to build AI enabled cyber traps and decoys

The Cyber Security Cooperative Research Centre (CSCRC), has today announced a major research project between homegrown cyber security company Penten and CSIRO's Data61, the data and digital specialist arm of Australia's national science agency, to extend the country's sovereign advantage in autonomous and active defence.

Applied Research -TRL


Research & Development
Field Validation & Demonstration
Scale-Up & Commercial Launch
Deployment
Market Transformation
Basic concept proven in laboratory setting
Product prototyping Pilot production
Expand manufacturing and supply chain, Protect IP, Spin-out or license
First commercial sales, Early codes and standards
Robust commercial sales, Product iterations and improvements, Established codes and standards
LEVEL OF INVOLVEMENT
Technological Valley of Death
Commercialization Valley of Death
Profitability Valley of Death
Investors
Banks
Government
Industry
TRL 1-4
TRL 5-6
TRL 7-8
TRL 9
MRL 10-12
MRL 13-15
Technology Scaling
Research & Development (R&D)
Market Scaling
TECHNOLOGY RISK
TRL = Technology Readiness Level
MRL = Market Readiness Level

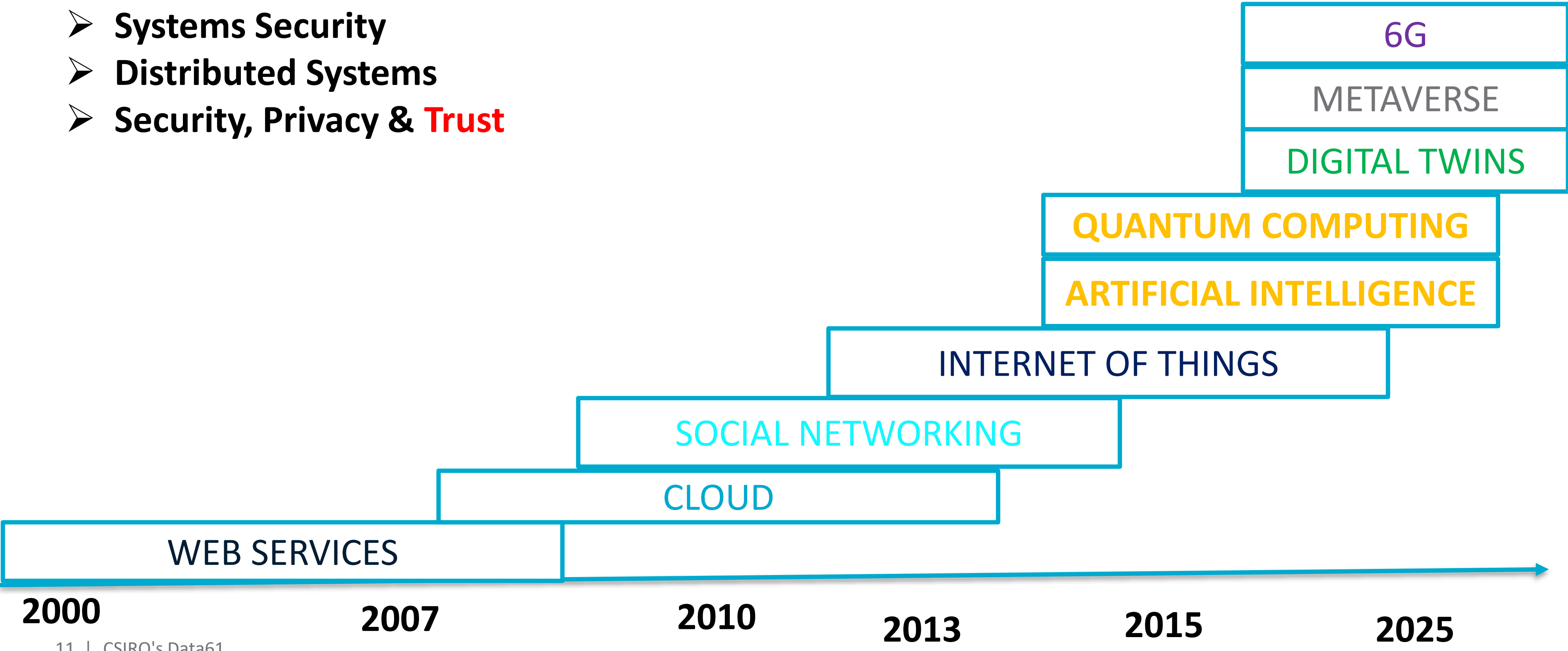
Applied Research – IRL



Investment Readiness Level

Identify and Validate Metrics That Matter	IRL 9
Validate Value Delivery (Left side of Canvas)	IRL 8
Prototype High-Fidelity Min. Viable Product	IRL 7
Validate Revenue Model (Right side of Canvas)	IRL 6
Validate Product/Market Fit	IRL 5
Prototype Low-Fidelity Min. Viable Product	IRL 4
Problem / Solution Validation	IRL 3
Market Size/Competitive Analysis	IRL 2
Complete First-Pass Business Model Canvas	IRL 1

- **Cybersecurity & Emerging Technologies**
- **Systems Security**
- **Distributed Systems**
- **Security, Privacy & Trust**



6G

METAVERSE

DIGITAL TWINS

QUANTUM COMPUTING

ARTIFICIAL INTELLIGENCE

INTERNET OF THINGS

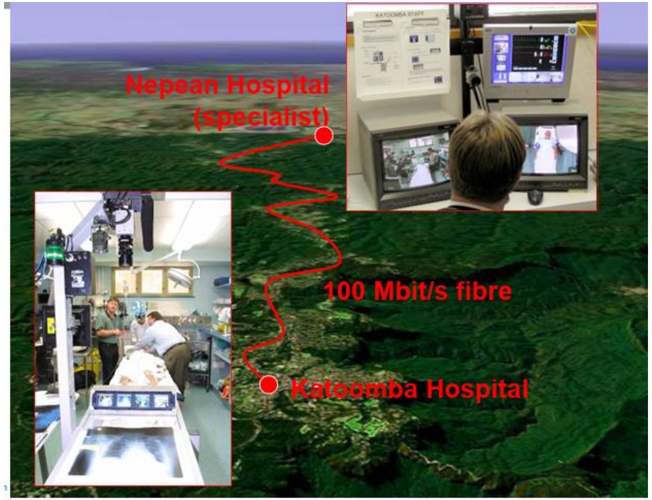
SOCIAL NETWORKING

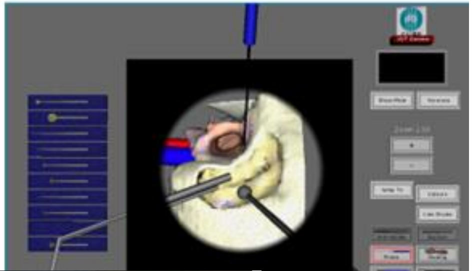
CLOUD

WEB SERVICES

Centre for Networking Technologies for the Information Economy (CeNTIE)

- Established and led by Terry Percival (one of the WiFi inventors)
- Started in 2001 with A$14M government funding and a total of $44M
- CeNTIE rolled out a prototype national broadband network connecting 18 nodes from Sydney to Canberra, Melbourne and Perth at 1 Gbit/s or higher.
- Number of target applications
  - **creation of collaborative networks for the film post-production industry**
  - virtual reality surgical training,
  - distance education and
  - tele-health
- A$10M extension funding



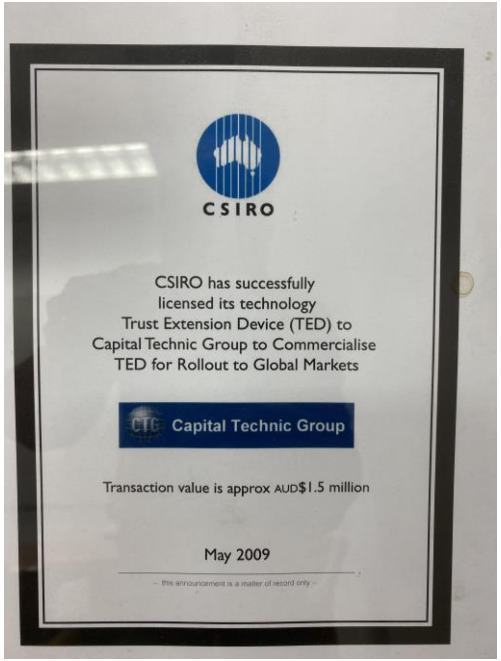


Telehealth platform - Coviu

Data61 spin-out Coviu is a telehealth platform that allows all clinicians to connect to their patients remotely. Practitioners of all professions can set up their own digital practice in under five minutes and start delivering end-to-end encrypted services immediately. Since mid-March 2020 the COVID-19 pandemic has seen a rapid uptake in Australian healthcare businesses employing Coviu - with now over 10,000 medical professionals using the platform to provide comprehensive, safe, and quarantine-compliant healthcare to their patients. Coviu was spun out of Data61 in May 2018 with venture funding from the CSIRO Innovation Fund managed by Main Sequence Ventures.

Woman using the Coviu platform to display a medical image via a computer.

Trust Extension Device



CSIRO

CSIRO has successfully licensed its technology Trust Extension Device (TED) to Capital Technic Group to Commercialise TED for Rollout to Global Markets

CTG Capital Technic Group

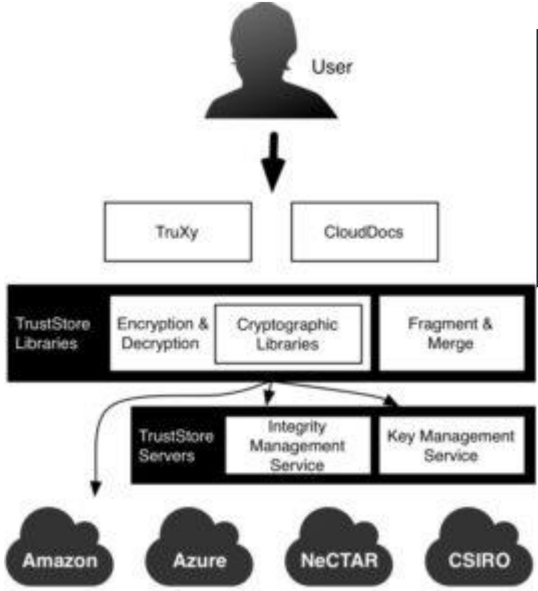
Transaction value is approx AUD$1.5 million

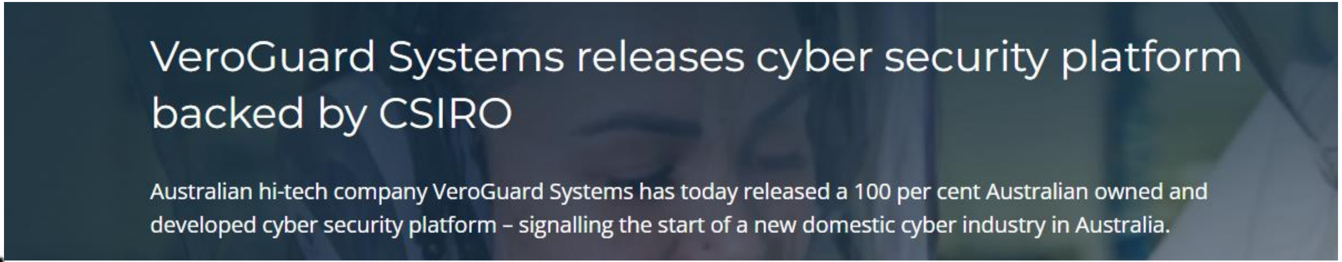
May 2009

– this announcement is a matter of record only –

John Zic
Surya Nepal
Dongxi Liu
.....

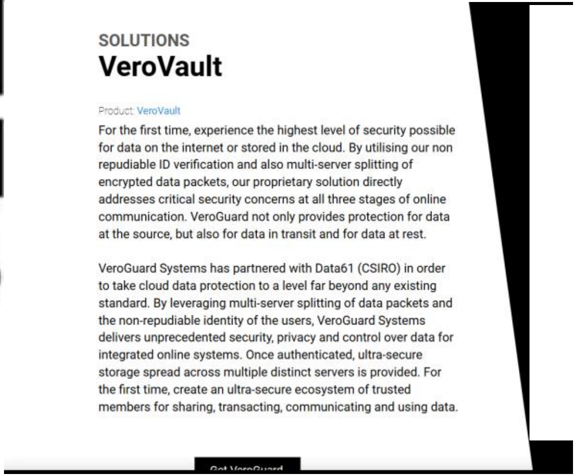
Secure Distributed Storage - TrustStore





VeroGuard Systems releases cyber security platform backed by CSIRO

Australian hi-tech company VeroGuard Systems has today released a 100 per cent Australian owned and developed cyber security platform – signalling the start of a new domestic cyber industry in Australia.



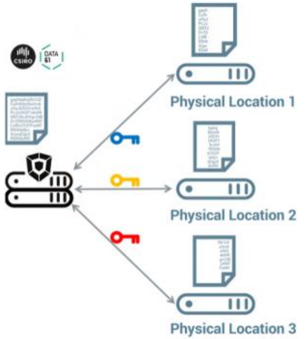
SOLUTIONS

VeroVault

Product: VeroVault

For the first time, experience the highest level of security possible for data on the internet or stored in the cloud. By utilising our non repudiable ID verification and also multi-server splitting of encrypted data packets, our proprietary solution directly addresses critical security concerns at all three stages of online communication. VeroGuard not only provides protection for data at the source, but also for data in transit and for data at rest.

VeroGuard Systems has partnered with Data61 (CSIRO) in order to take cloud data protection to a level far beyond any existing standard. By leveraging multi-server splitting of data packets and the non-repudiable identity of the users, VeroGuard Systems delivers unprecedented security, privacy and control over data for integrated online systems. Once authenticated, ultra-secure storage spread across multiple distinct servers is provided. For the first time, create an ultra-secure ecosystem of trusted members for sharing, transacting, communicating and using data.

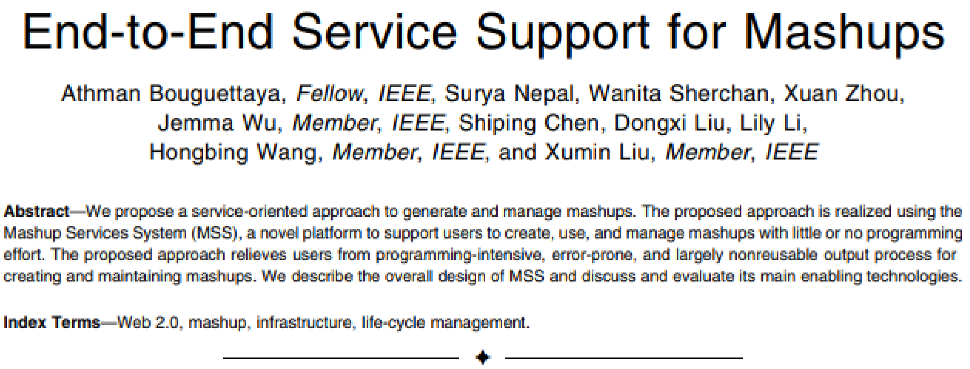


Paul Greenfield
Paul Watters
Shiping Chen
John Zic
Surya Nepal
.....

Human Services Delivery Research Alliance

- Established in 2009
- A five-year research alliance between the CSIRO and Centrelink, committing $25 million to drive a significant program of improving Australian Government service delivery.
- The focus was on **Trust**

Trust Management in Services



End-to-End Service Support for Mashups

Athman Bouguettaya, *Fellow, IEEE*, Surya Nepal, Wanita Sherchan, Xuan Zhou,
Jemma Wu, *Member, IEEE*, Shiping Chen, Dongxi Liu, Lily Li,
Hongbing Wang, *Member, IEEE*, and Xumin Liu, *Member, IEEE*

Abstract—We propose a service-oriented approach to generate and manage mashups. The proposed approach is realized using the Mashup Services System (MSS), a novel platform to support users to create, use, and manage mashups with little or no programming effort. The proposed approach relieves users from programming-intensive, error-prone, and largely nonreusable output process for creating and maintaining mashups. We describe the overall design of MSS and discuss and evaluate its main enabling technologies.

Index Terms—Web 2.0, mashup, infrastructure, life-cycle management.

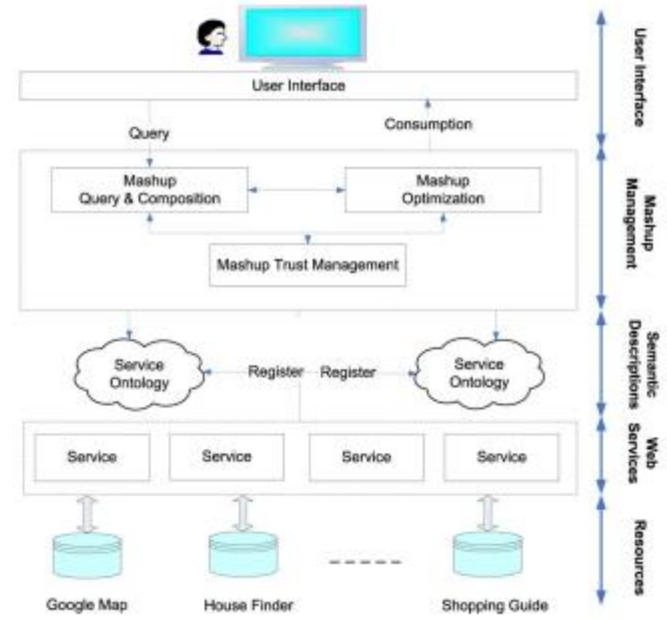
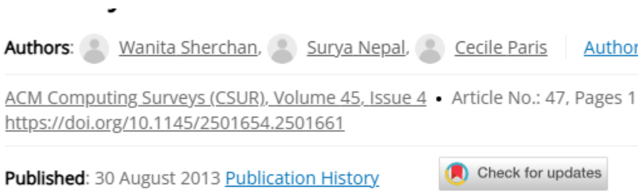


Fig. 1. Mashup services system architecture.

Trust in Social Networks

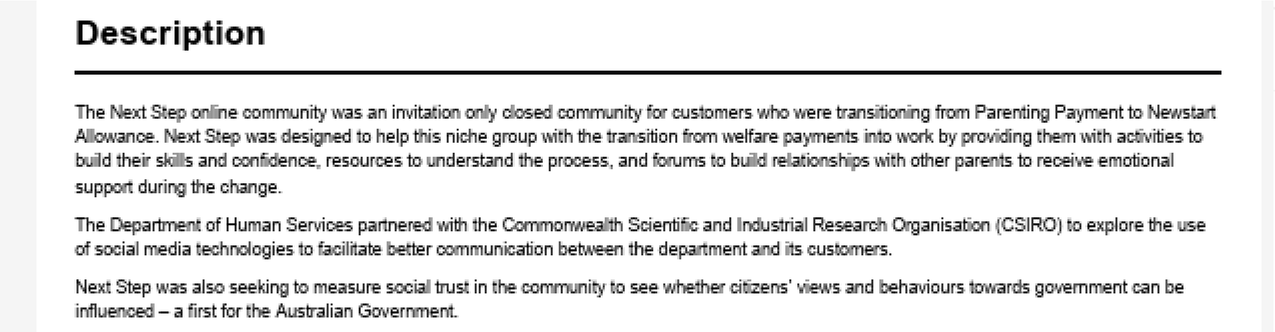


Authors: Wanita Sherchan, Surya Nepal, Cecile Paris Author

ACM Computing Surveys (CSUR), Volume 45, Issue 4 • Article No.: 47, Pages 1

https://doi.org/10.1145/2501654.2501661

Published: 30 August 2013 Publication History

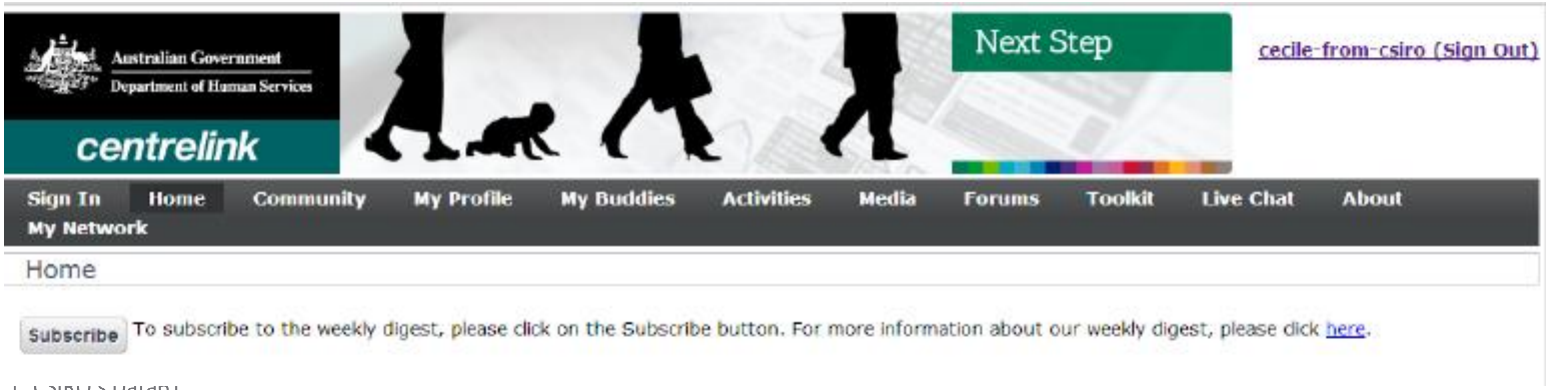


Description

The Next Step online community was an invitation only closed community for customers who were transitioning from Parenting Payment to Newstart Allowance. Next Step was designed to help this niche group with the transition from welfare payments into work by providing them with activities to build their skills and confidence, resources to understand the process, and forums to build relationships with other parents to receive emotional support during the change.

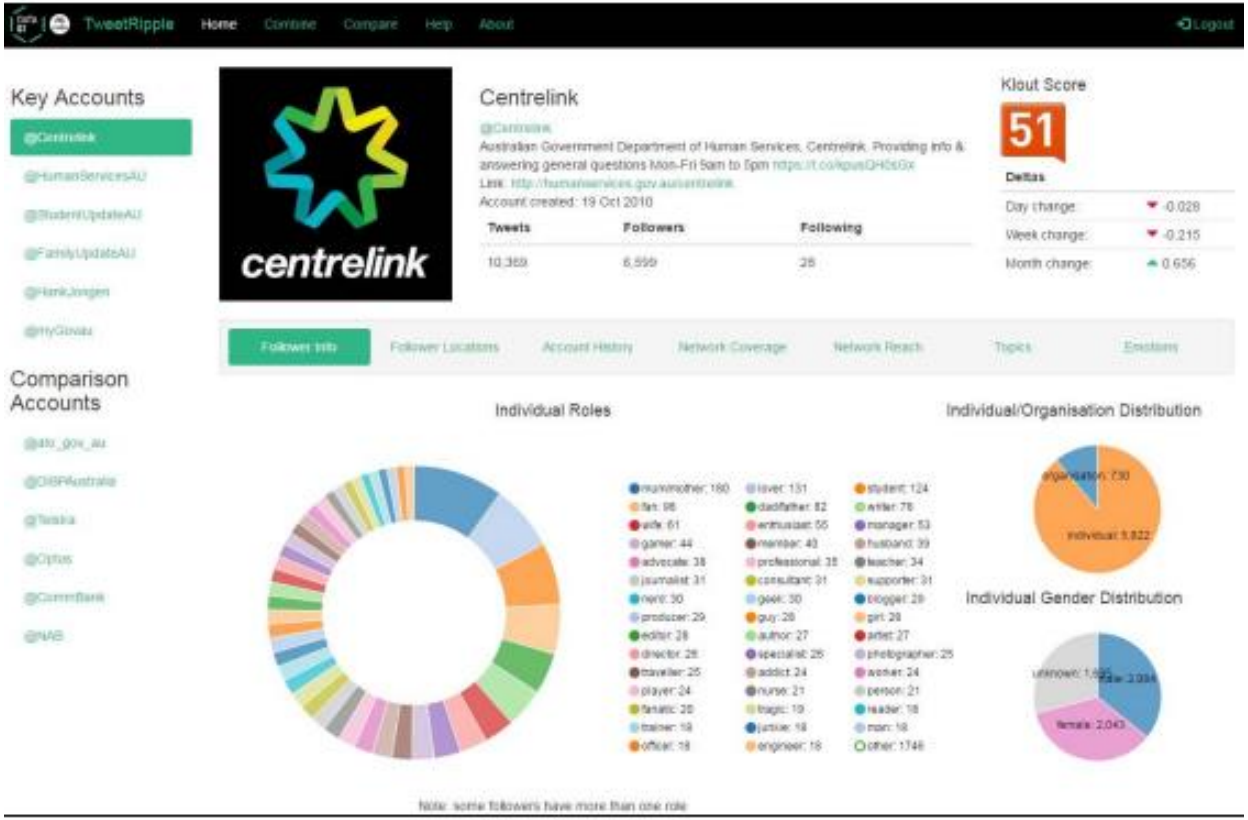
The Department of Human Services partnered with the Commonwealth Scientific and Industrial Research Organisation (CSIRO) to explore the use of social media technologies to facilitate better communication between the department and its customers.

Next Step was also seeking to measure social trust in the community to see whether citizens' views and behaviours towards government can be influenced – a first for the Australian Government.



To subscribe to the weekly digest, please click on the Subscribe button. For more information about our weekly digest, please click here.

Tweet Ripple



Centrelink

@Centrelink

Australian Government Department of Human Services, Centrelink. Providing info & answering general questions Mon-Fri 9am to 5pm https://t.co/kpusQH6sGx

Link: http://humanservices.gov.au/centrelink

Account created: 19 Oct 2010

Tweets	Followers	Following
10,369	6,599	28

Klout Score

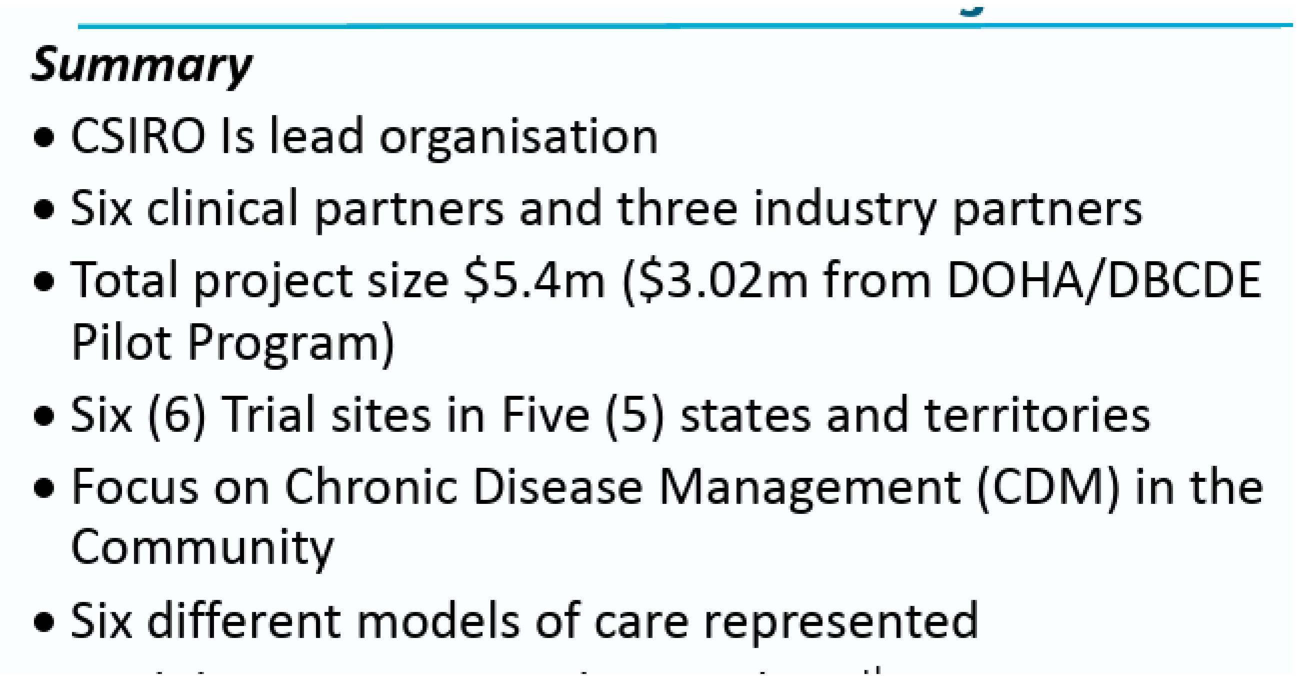
51

Deltas	
Day change:	▼ -0.028
Week change:	▼ -0.215
Month change:	▲ 0.656

Note: some followers have more than one role

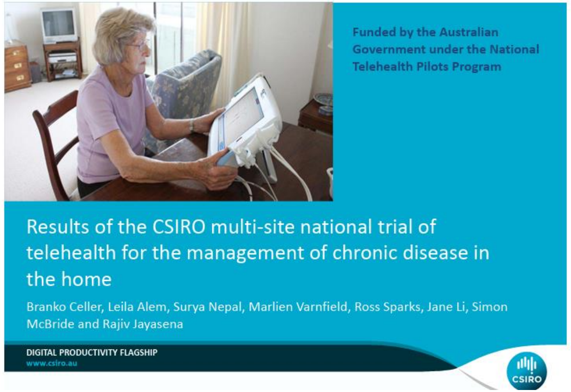
Fig 2: TweetRipple Web Application

NBN Telehealth



Summary

- CSIRO Is lead organisation
- Six clinical partners and three industry partners
- Total project size $5.4m ($3.02m from DOHA/DBCDE Pilot Program)
- Six (6) Trial sites in Five (5) states and territories
- Focus on Chronic Disease Management (CDM) in the Community
- Six different models of care represented



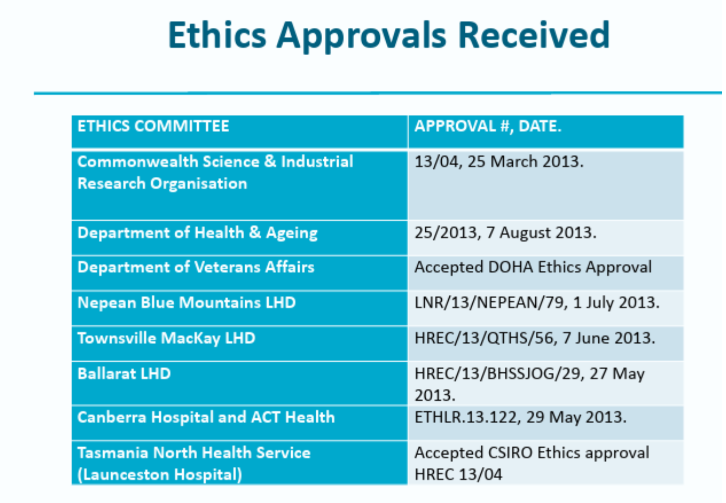
Funded by the Australian Government under the National Telehealth Pilots Program

Results of the CSIRO multi-site national trial of telehealth for the management of chronic disease in the home

Branko Celler, Leila Alem, Surya Nepal, Marlien Varnfield, Ross Sparks, Jane Li, Simon McBride and Rajiv Jayasena

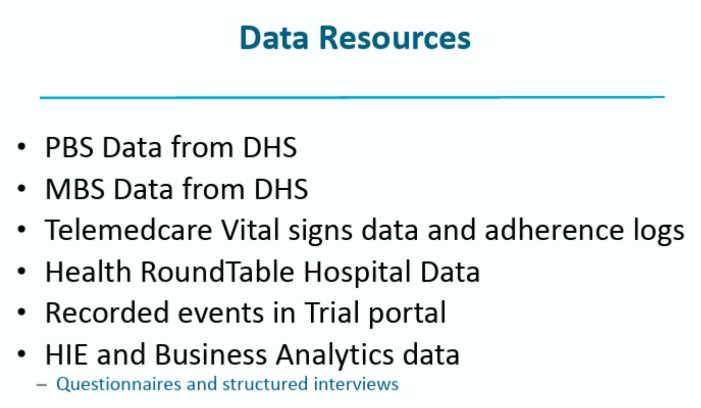
DIGITAL PRODUCTIVITY FLAGSHIP
www.csiro.au

Secure Data Management



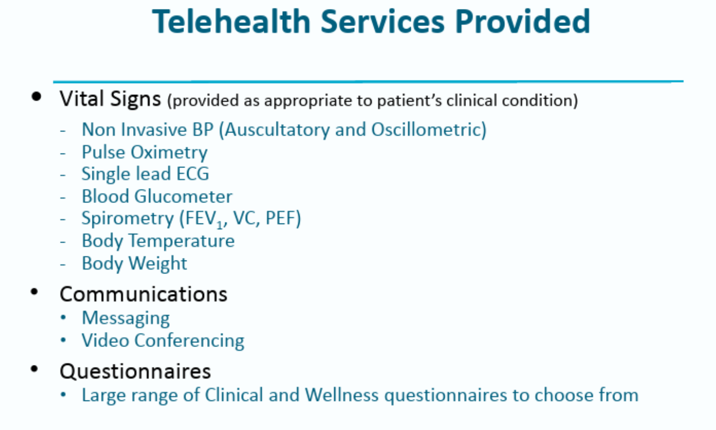
Ethics Approvals Received

ETHICS COMMITTEE	APPROVAL #, DATE.
Commonwealth Science & Industrial Research Organisation	13/04, 25 March 2013.
Department of Health & Ageing	25/2013, 7 August 2013.
Department of Veterans Affairs	Accepted DOHA Ethics Approval
Nepean Blue Mountains LHD	LNR/13/NEPEAN/79, 1 July 2013.
Townsville MacKay LHD	HREC/13/QTHS/56, 7 June 2013.
Ballarat LHD	HREC/13/BHSSJOG/29, 27 May 2013.
Canberra Hospital and ACT Health	ETHLR.13.122, 29 May 2013.
Tasmania North Health Service (Launceston Hospital)	Accepted CSIRO Ethics approval HREC 13/04



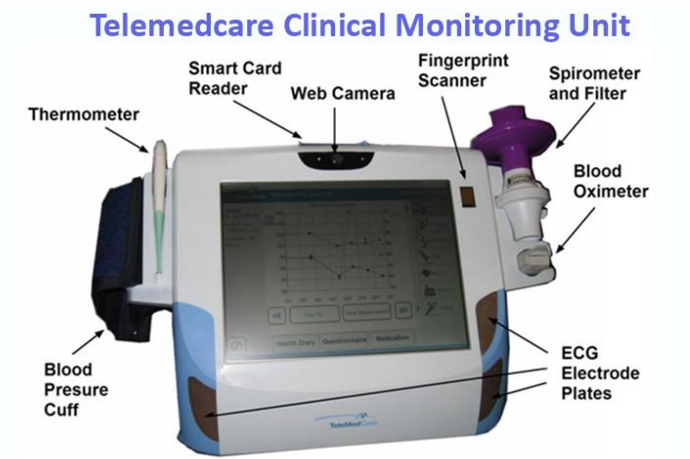
Data Resources

- PBS Data from DHS
- MBS Data from DHS
- Telemedcare Vital signs data and adherence logs
- Health RoundTable Hospital Data
- Recorded events in Trial portal
- HIE and Business Analytics data
  - Questionnaires and structured interviews



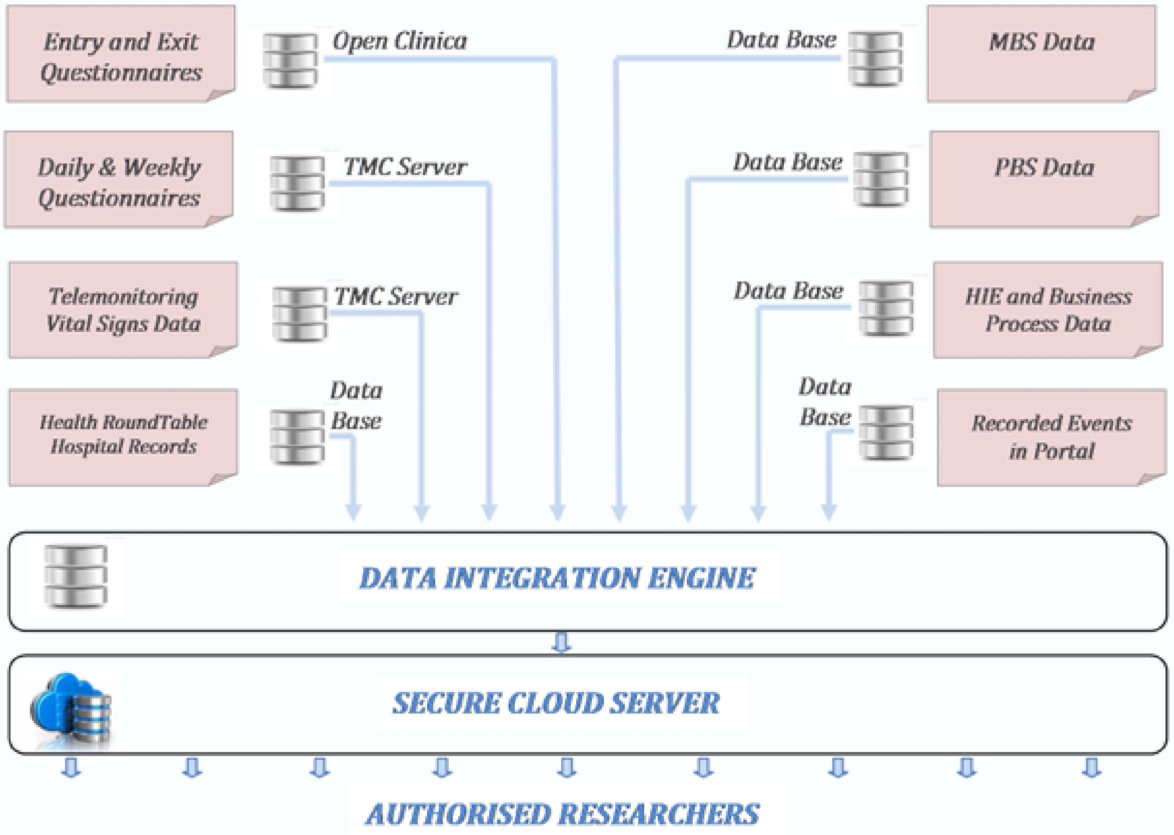
Telehealth Services Provided

- Vital Signs (provided as appropriate to patient's clinical condition)
  - Non Invasive BP (Auscultatory and Oscillometric)
  - Pulse Oximetry
  - Single lead ECG
  - Blood Glucometer
  - Spirometry (FEV$_1$, VC, PEF)
  - Body Temperature
  - Body Weight
- Communications
  - Messaging
  - Video Conferencing
- Questionnaires
  - Large range of Clinical and Wellness questionnaires to choose from



Telemedcare Clinical Monitoring Unit

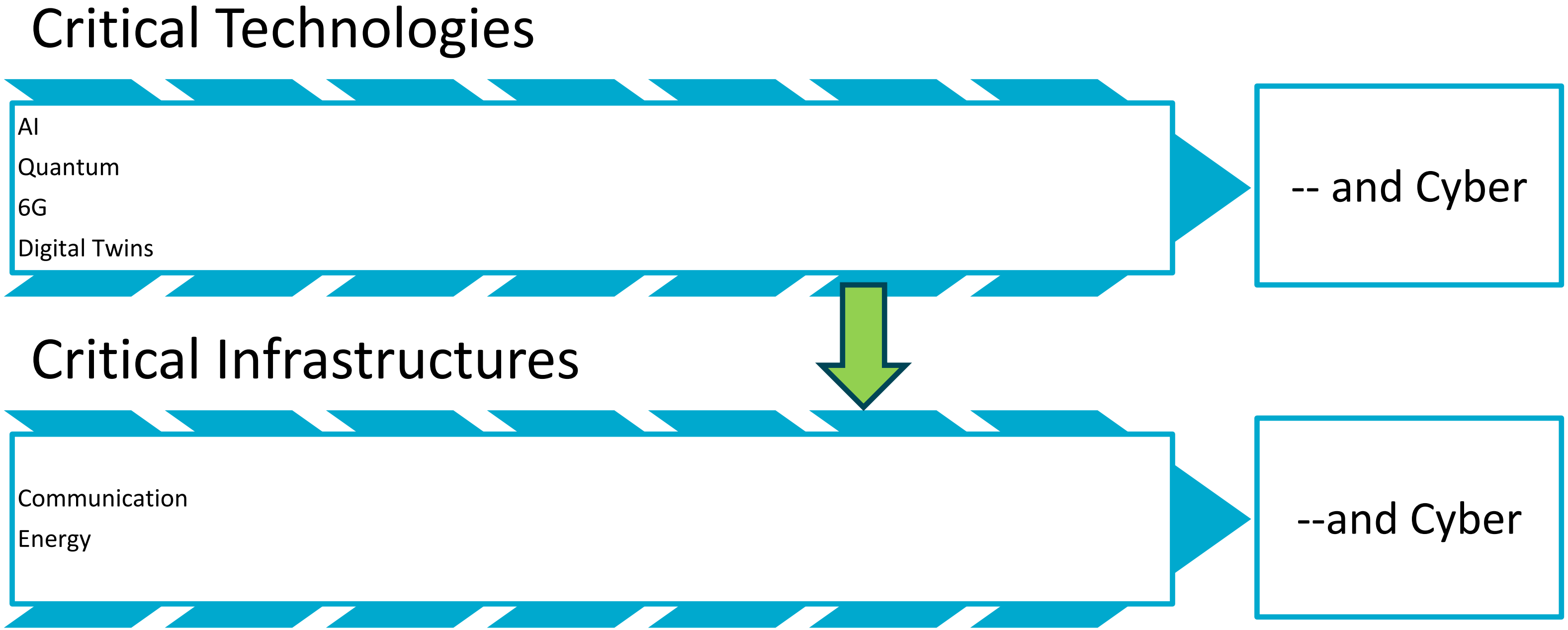
Integration of multiple data sources


Entry and Exit Questionnaires
Open Clinica
Daily & Weekly Questionnaires
TMC Server
Telemonitoring Vital Signs Data
TMC Server
Health RoundTable Hospital Records
Data Base
Data Base
MBS Data
Data Base
PBS Data
Data Base
HIE and Business Process Data
Data Base
Recorded Events in Portal
DATA INTEGRATION ENGINE
SECURE CLOUD SERVER
AUTHORISED RESEARCHERS

Data61

Australia's National Science Agency

Critical Technologies and Critical Infrastructure



Critical Technologies

AI

Quantum

6G

Digital Twins

-- and Cyber

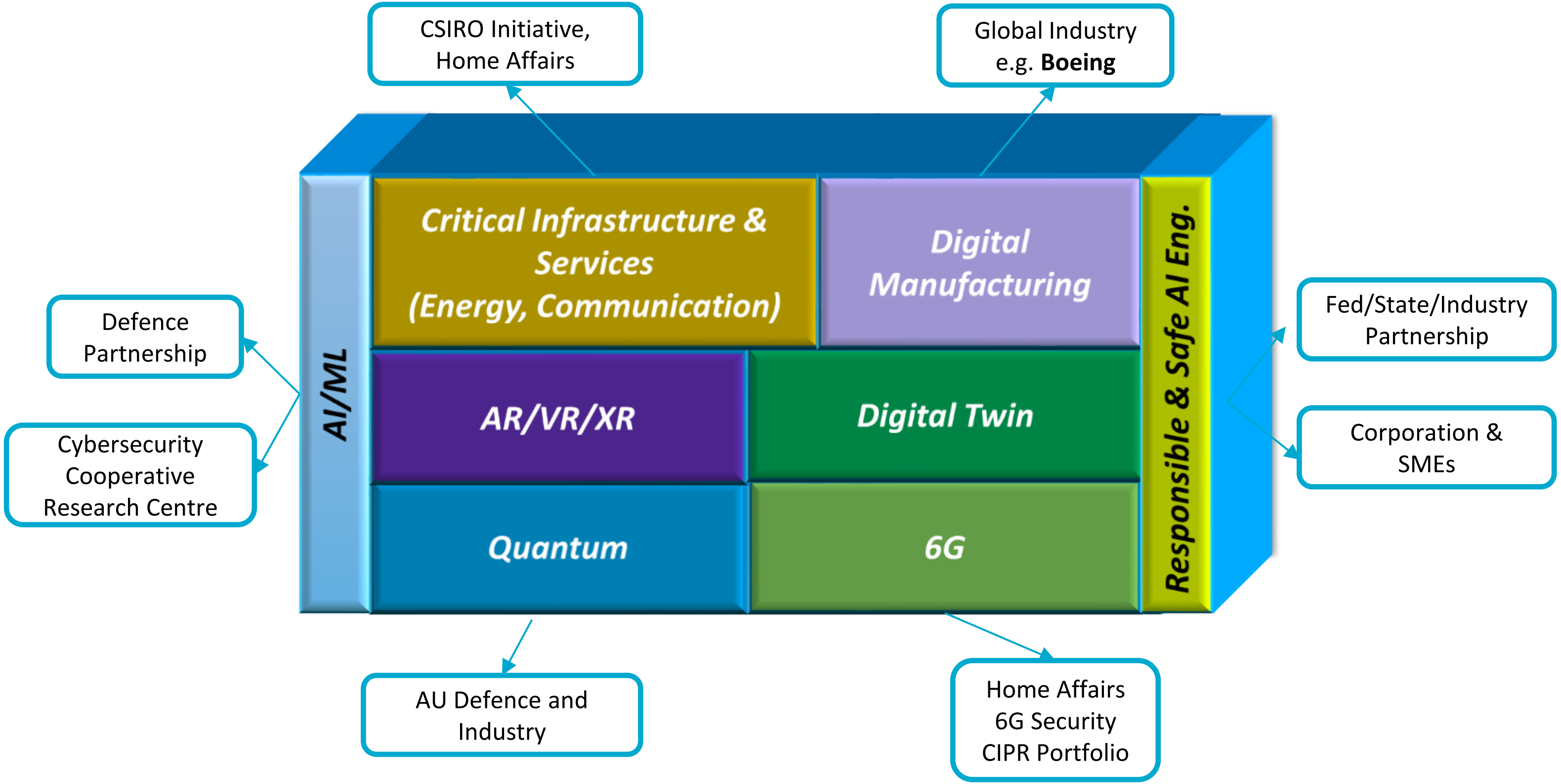
Critical Infrastructures

Communication

Energy

--and Cyber

Future S&T Stack and Partners



CSIRO Initiative, Home Affairs

Global Industry e.g. **Boeing**

Defence Partnership

Cybersecurity Cooperative Research Centre

Fed/State/Industry Partnership

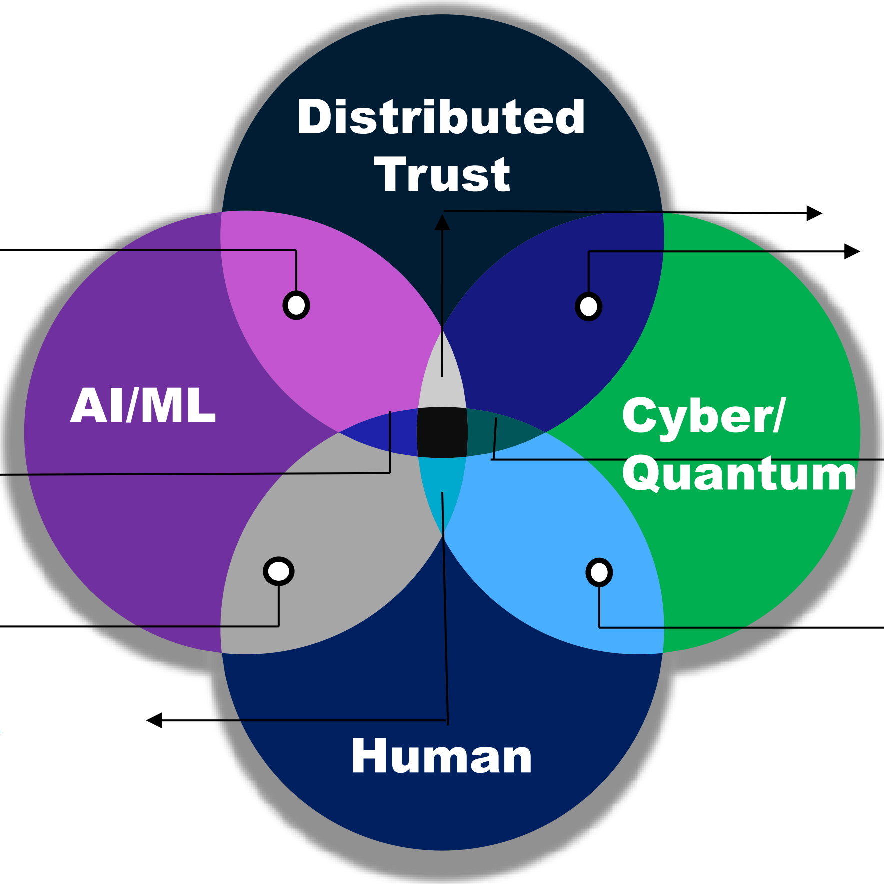
Corporation & SMEs

AU Defence and Industry

Home Affairs
6G Security
CIPR Portfolio

AI/ML
Critical Infrastructure & Services (Energy, Communication)
Digital Manufacturing
AR/VR/XR
Digital Twin
Quantum
6G
Responsible & Safe AI Eng.

Combinatorial Innovation



Trustworthy/Federated AI/Agents & Quantum-Safe Quantum ML

Trustworthy AI/Data Ecosystem

Trustworthy Digital Infrastructure

Trustworthy AI Practices & Supply Chains

Sovereign Identity & Transaction

Responsible AI

Human-Centred Cybersecurity

Trusted Collaborative Intelligence

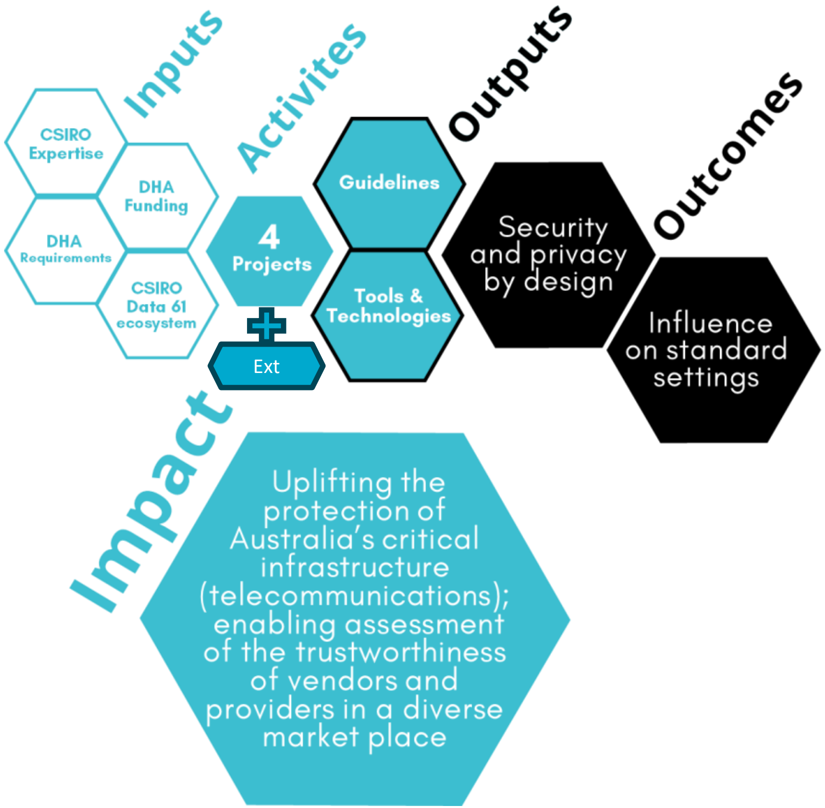
Critical Infrastructure

Communication

Energy

Australia’s National Science Agency

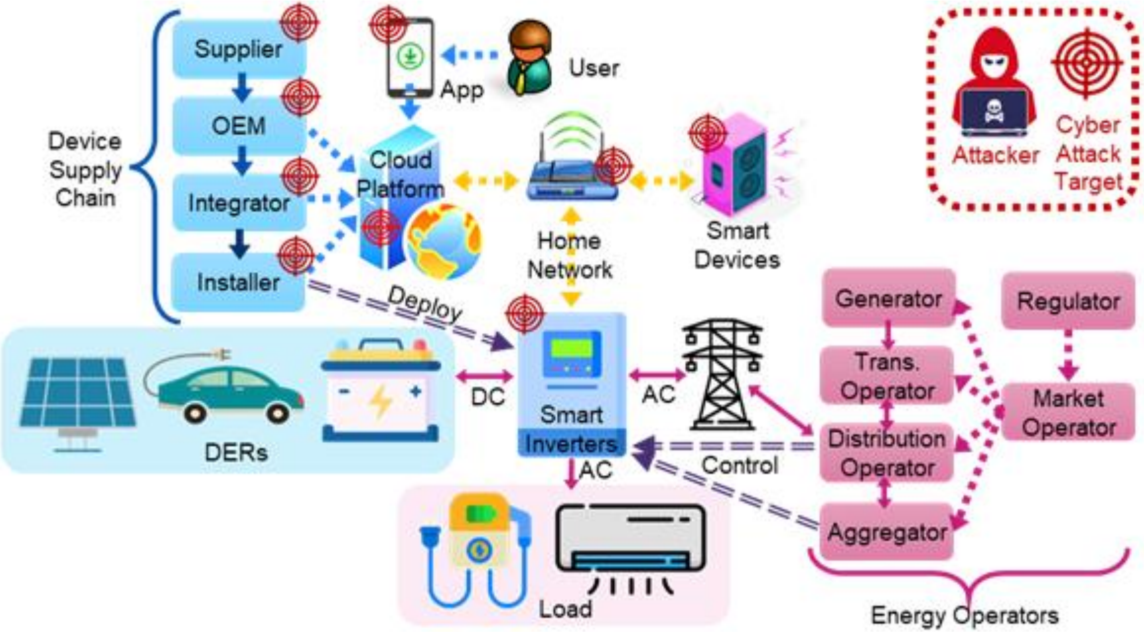
6G Security Research and Development Program


Inputs
CSIRO Expertise
DHA Funding
DHA Requirements
CSIRO Data 61 ecosystem
Activites
4 Projects
Ext
Outputs
Guidelines
Tools & Technologies
Security and privacy by design
Outcomes
Influence on standard settings
Impact
Uplifting the protection of Australia's critical infrastructure (telecommunications); enabling assessment of the trustworthiness of vendors and providers in a diverse market place

foundational research into the security requirements of 6G and future connectivity technologies, ensuring they are secure-by-design and help shape international standards in a way that aligns with our values and expectations around security

Partner: Department of Home Affairs (A$12.25M)

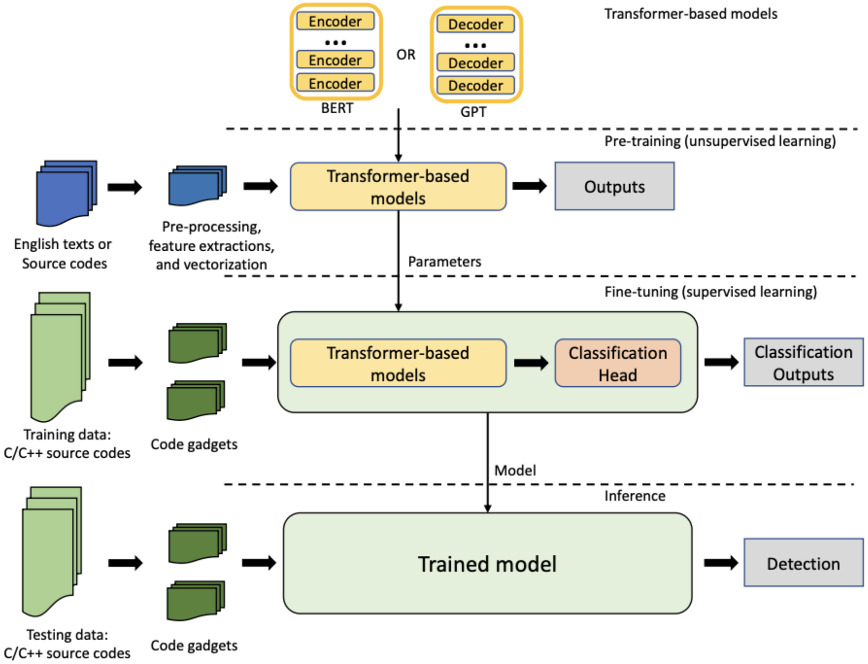
Complex Ecosystem, Intricate Attack Vectors in Energy ecosystem

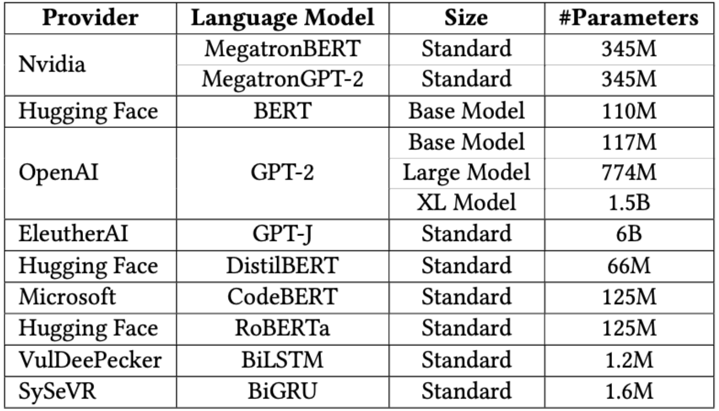


AI & Cyber

Australia's National Science Agency

Source code vulnerability detection





Provider	Language Model	Size	#Parameters
Nvidia	MegatronBERT	Standard	345M
	MegatronGPT-2	Standard	345M
Hugging Face	BERT	Base Model	110M
OpenAI	GPT-2	Base Model	117M
		Large Model	774M
		XL Model	1.5B
EleutherAI	GPT-J	Standard	6B
Hugging Face	DistilBERT	Standard	66M
Microsoft	CodeBERT	Standard	125M
Hugging Face	RoBERTa	Standard	125M
VulDeePecker	BiLSTM	Standard	1.2M
SySeVR	BiGRU	Standard	1.6M

Source: https://dl.acm.org/doi/pdf/10.1145/3564625.3567985

Ref: Chandra Thapa, Seung Ick Jang, Muhammad Ejaz Ahmed, Seyit Camtepe, Josef Pieprzyk, Surya Nepal, "*Transformer-based language models for software vulnerability detection*," ACSAC, 2022.

Data61 AI & Cybersecurity Research

Security and Safety of AI Systems

- Integrity of AI Models
- Red Teaming and adversarial testing
- Poisoning and backdooring
- Machine unlearning

Mitigating AI Risk for Secure/Safe Adoption

- Synthetic Content (Deepfake Misinformation, software)
- Synthetic Actions (Agentic AI)
- Synthetic info & knowledge
- Tasks Automation & Amplified Risk at Scale

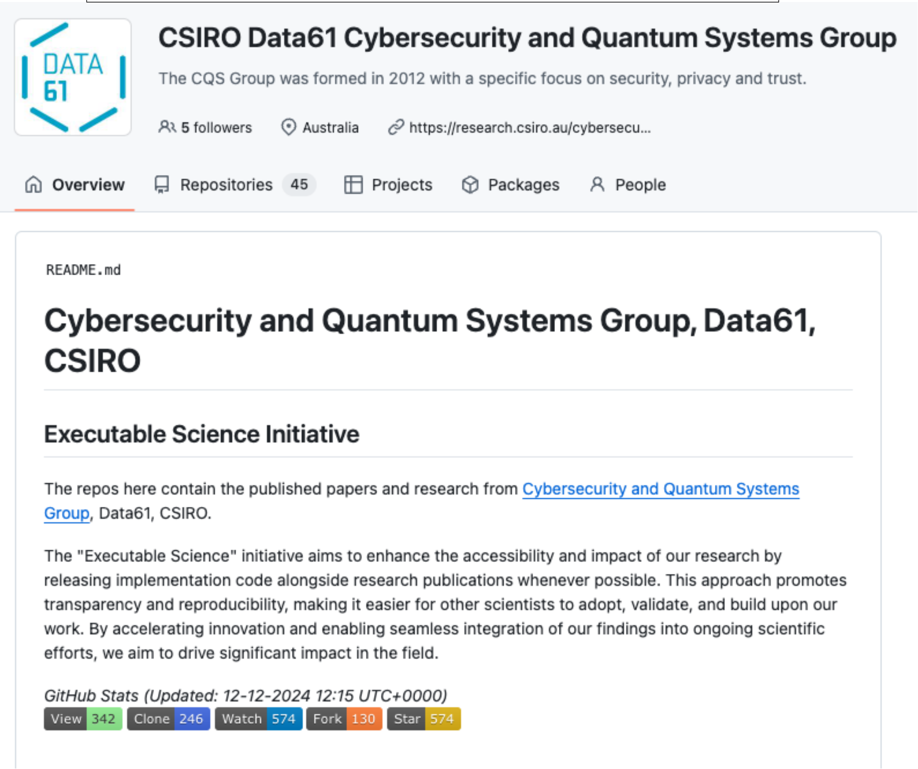
Applications of AI for Cybersecurity

- NLP for Cyber (Malware, Phishing, Ransomware, Vulnerabilities)
- Active Cyber Defence
- Deception technologies
- Human-AI teaming

Enabling Secure and Safe AI Adoption to Drive National Productivity and Competitiveness

Online Library

https://github.com/cqs-data61



CSIRO Data61 Cybersecurity and Quantum Systems Group

The CQS Group was formed in 2012 with a specific focus on security, privacy and trust.

5 followers   Australia   https://research.csiro.au/cybersecu...

README.md

Cybersecurity and Quantum Systems Group, Data61, CSIRO

Executable Science Initiative

The repos here contain the published papers and research from Cybersecurity and Quantum Systems Group, Data61, CSIRO.

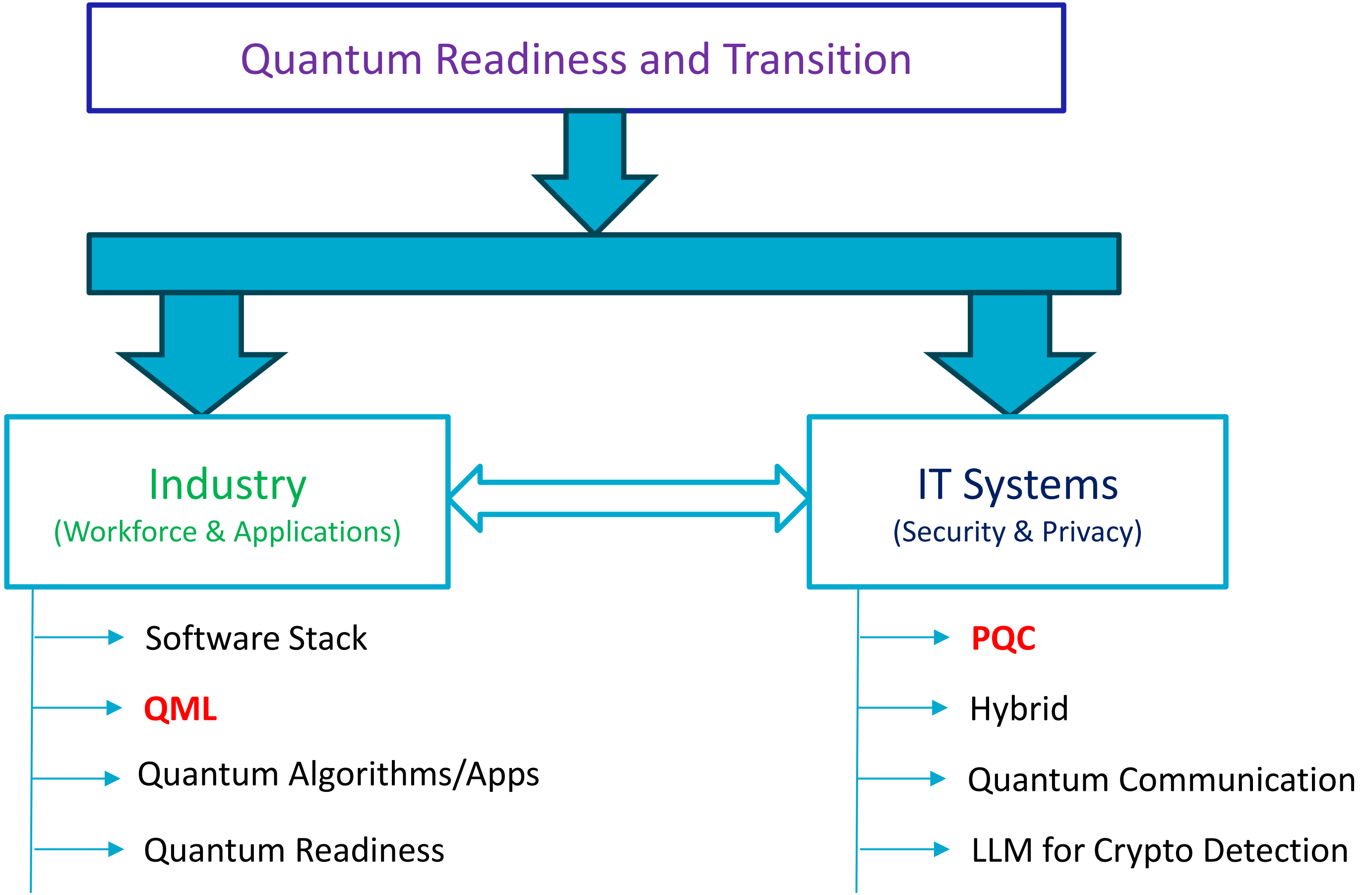
The "Executable Science" initiative aims to enhance the accessibility and impact of our research by releasing implementation code alongside research publications whenever possible. This approach promotes transparency and reproducibility, making it easier for other scientists to adopt, validate, and build upon our work. By accelerating innovation and enabling seamless integration of our findings into ongoing scientific efforts, we aim to drive significant impact in the field.

GitHub Stats (Updated: 12-12-2024 12:15 UTC+0000)

View 342   Clone 246   Watch 574   Fork 130   Star 574

Quantum: Opportunities and Threats

Australia’s National Science Agency



Quantum Readiness and Transition

Industry

(Workforce & Applications)

- Software Stack
- **QML**
- Quantum Algorithms/Apps
- Quantum Readiness

IT Systems

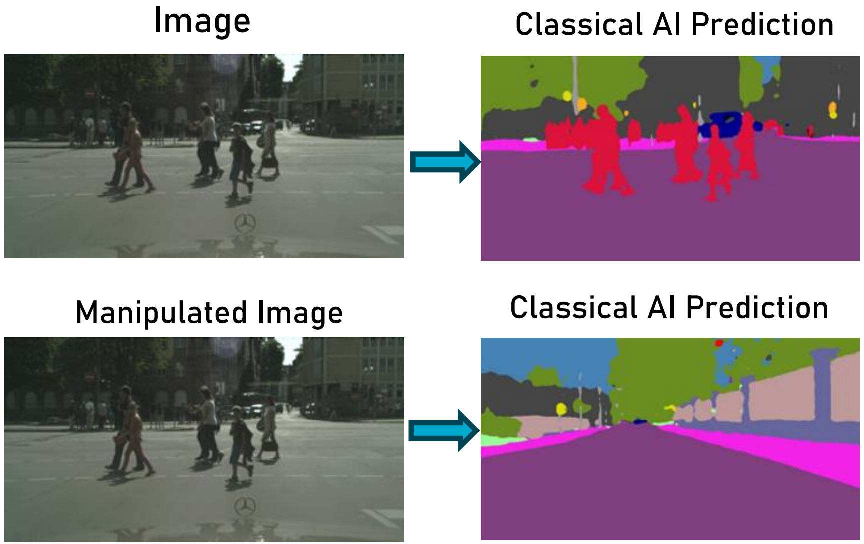
(Security & Privacy)

- **PQC**
- Hybrid
- Quantum Communication
- LLM for Crypto Detection

Quantum AI – AU Army and ASCA (Over A$4M)

Artificial Intelligence Algorithms are widely used in security-sensitive applications involving images and signals:

However, artificial intelligence vulnerabilities are a big threat.



Image

Classical AI Prediction

Manipulated Image

Classical AI Prediction

J. Metzen et al., ICCV paper, Computer Vision Foundation

Current Classical Solutions rely on better training of Artificial Intelligence – Does **NOT** guarantee trust!

A transformative new technology is needed!

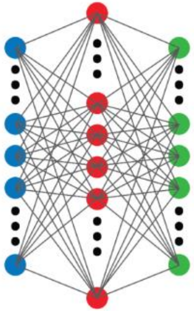
https://www.usni.org/magazines/proceedings/2021/november/prepare-ai-enabled-future-logistics

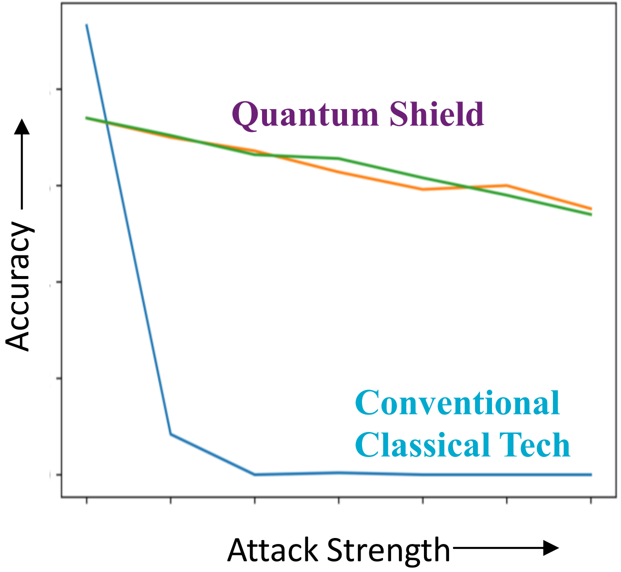
Quantum AI

Quantum Machine Learning is a **fundamentally new technology** working on the principles of quantum mechanics – superposition of dataset, entanglement between quantum neurons. It offers many advantages over classical counterparts, but most importantly for our purposes, it is **highly resilient** against adversarial and cyber attacks.

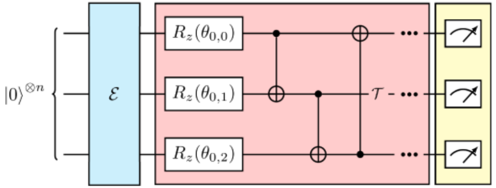
Datasets
Images or Signals

Conventional
Classical Technology

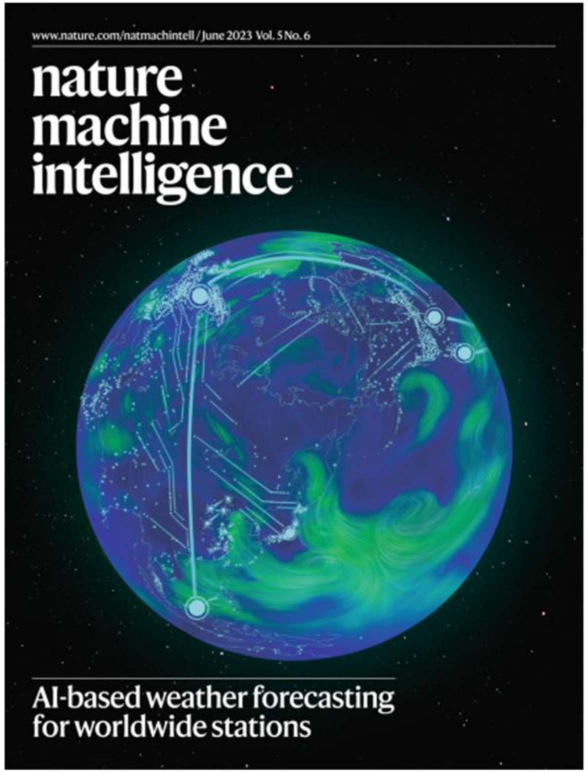




Quantum Shield for
Artificial Intelligence



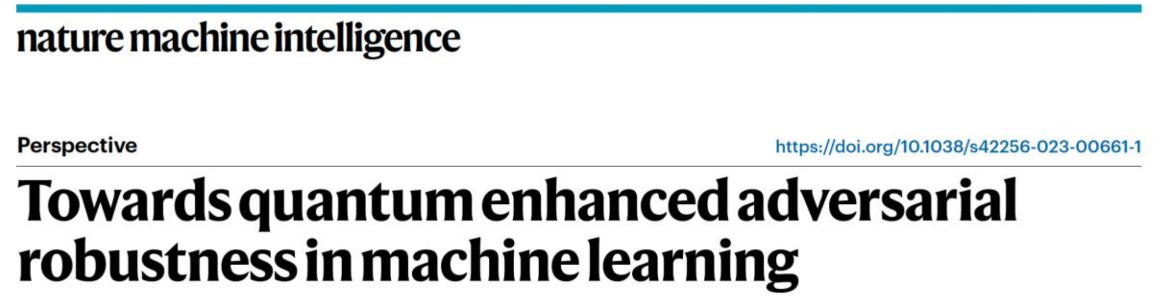
Our technology is world-leading



www.nature.com/natmachintell / June 2023 Vol. 5 No. 6

nature machine intelligence

AI-based weather forecasting for worldwide stations

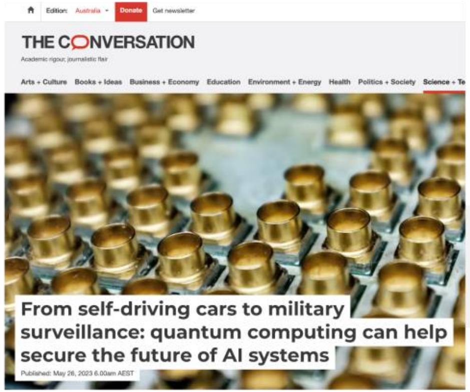


nature machine intelligence

Perspective

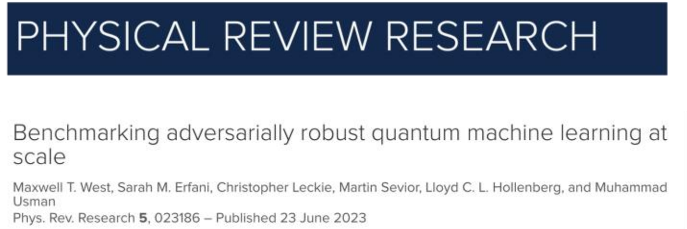
https://doi.org/10.1038/s42256-023-00661-1

Towards quantum enhanced adversarial robustness in machine learning



From self-driving cars to military surveillance: quantum computing can help secure the future of AI systems

Published: May 26, 2023 6.00am AEST



PHYSICAL REVIEW RESEARCH

Benchmarking adversarially robust quantum machine learning at scale

Maxwell T. West, Sarah M. Erfani, Christopher Leckie, Martin Sevior, Lloyd C. L. Hollenberg, and Muhammad Usman

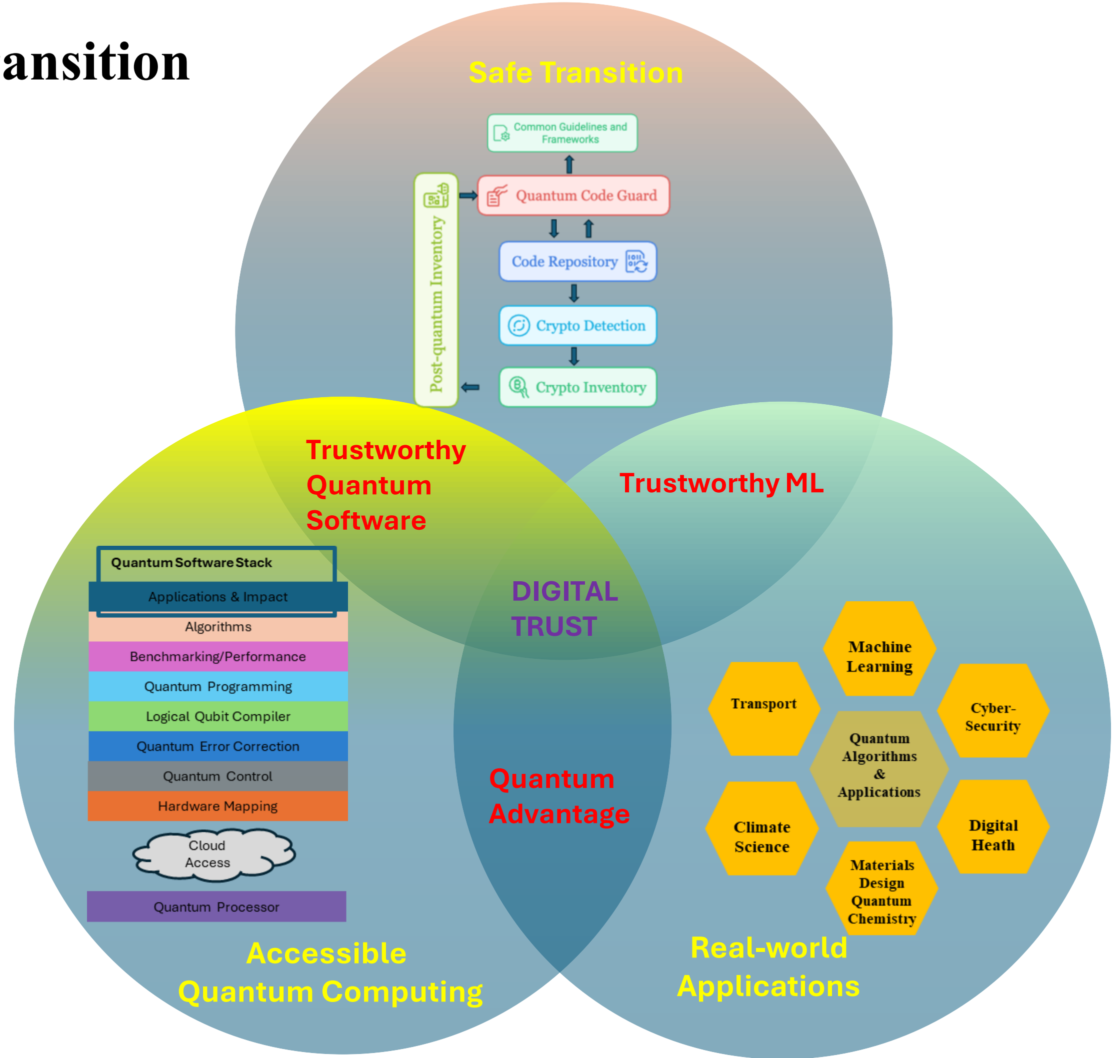
Phys. Rev. Research **5**, 023186 – Published 23 June 2023

Our work: accelerating transition with higher security

iAWARDS 25
Australian Capital Territory
GOVERNMENT & PUBLIC SECTOR
MIKA: A MINIMALIST APPROACH
TO HYBRID KEY EXCHANGE
CSIRO AND PENTEN PTY LTD

- High-assurance transition to quantum-safe VPN
  - MIKA: hybrid of pre-quantum and post-quantum VPN implementa
    - Avoiding introducing new vulnerabilities during transitions
    - Extensible to Quantum Key Distribution
  - **Collaboration with Penten**
- Quantum-safe 5G/6G protocols
  - Quantum-safe upgrading of OpenAirInterface 5G platform
  - A new quantum-safe mechanism to prevent Caller Spoofing, reducing scams in Australia
  - **Partial support from DHA**

Enabling Safe Transition



Safe Transition

Common Guidelines and Frameworks

Quantum Code Guard

Code Repository

Crypto Detection

Crypto Inventory

Post-quantum Inventory

Trustworthy Quantum Software

Trustworthy ML

DIGITAL TRUST

Quantum Software Stack

Applications & Impact

Algorithms

Benchmarking/Performance

Quantum Programming

Logical Qubit Compiler

Quantum Error Correction

Quantum Control

Hardware Mapping

Cloud Access

Quantum Processor

Quantum Advantage

Machine Learning

Transport

Cyber-Security

Quantum Algorithms & Applications

Climate Science

Digital Heath

Materials Design Quantum Chemistry

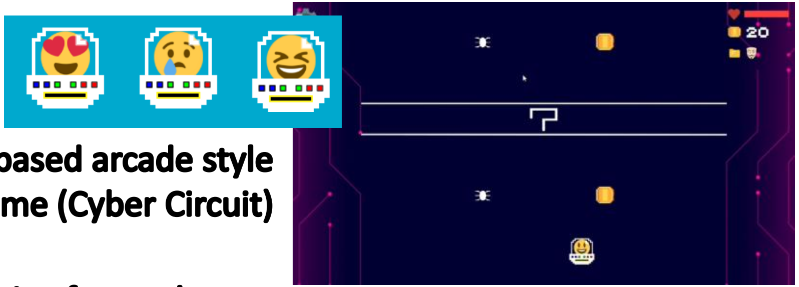
Accessible Quantum Computing

Real-world Applications

Human-Centric

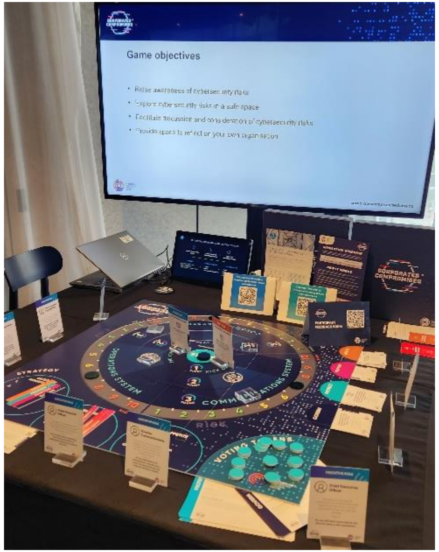
Cyber gamification

Table top event based cybersecurity game (CyberIQ)



Computer based arcade style game (Cyber Circuit)

Table top executive focused cybersecurity game (Corporates Compromised)



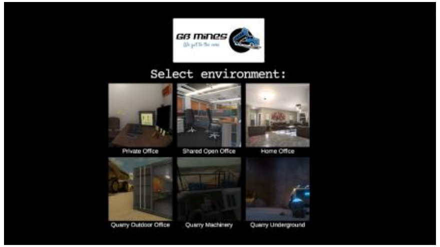
Game objectives

- Raise awareness of cybersecurity risks
- Explore cybersecurity risks in a safe space
- Facilitate discussion and consideration of cybersecurity risks
- Provide space to reflect on your own organisation

Immersive cybersecurity escape game (CyberSIM)

Continuous development platform (CRCounter)

Digital twin simulation



Select environment:

***For more information:** Marthie Grobler (marthie.grobler@data61.csiro.au)*

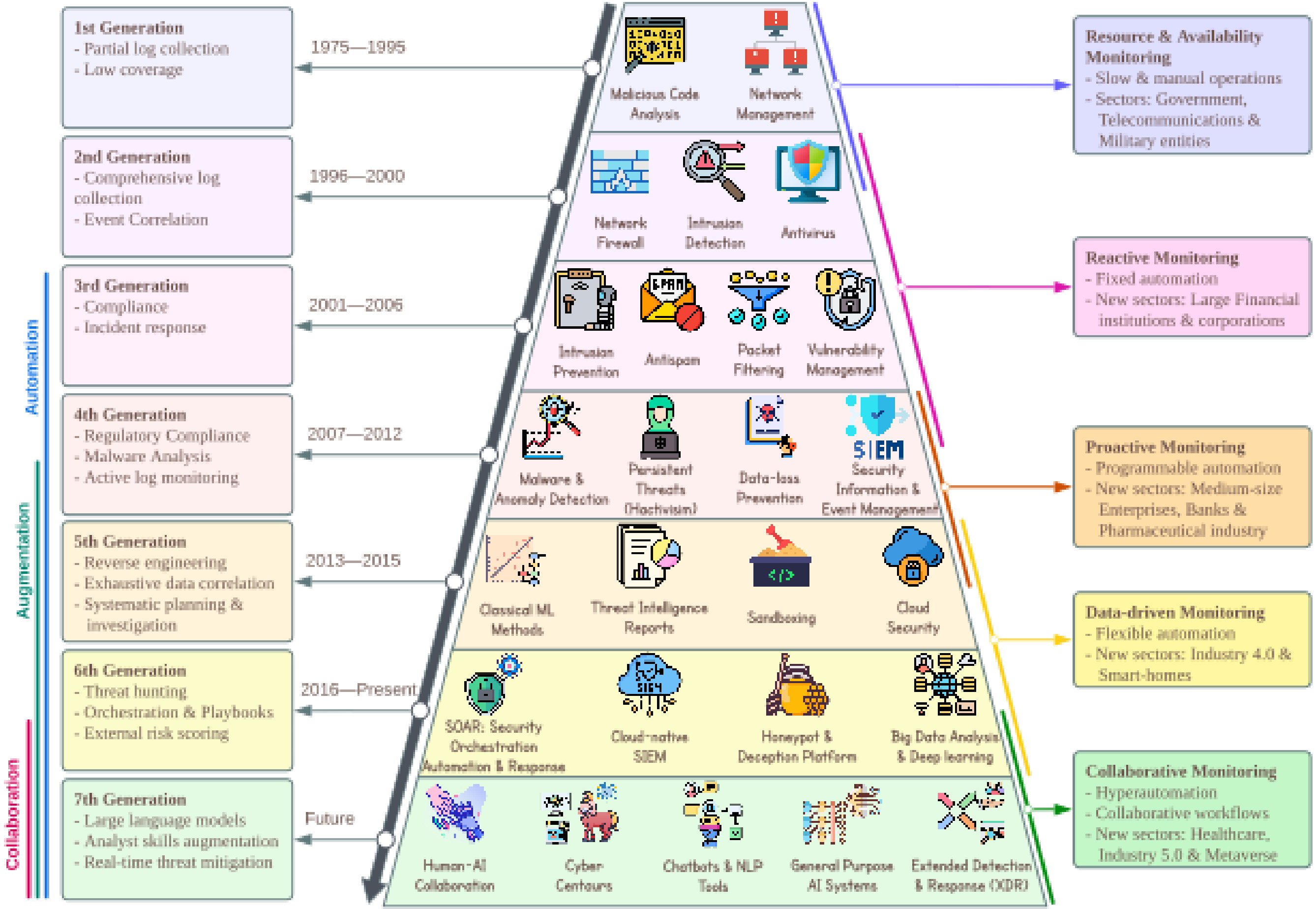
Evolution of SOCs

Traditional SOCs rely heavily on manual triage and rule-based systems.

Modern SOCs integrate automation and AI for threat detection.

Emerging trend: Human–AI collaboration to balance automation with human judgment.

Shift from reactive to proactive and adaptive security operations.



1st Generation
- Partial log collection
- Low coverage

1975—1995

2nd Generation
- Comprehensive log collection
- Event Correlation

1996—2000

3rd Generation
- Compliance
- Incident response

2001—2006

4th Generation
- Regulatory Compliance
- Malware Analysis
- Active log monitoring

2007—2012

5th Generation
- Reverse engineering
- Exhaustive data correlation
- Systematic planning & investigation

2013—2015

6th Generation
- Threat hunting
- Orchestration & Playbooks
- External risk scoring

2016—Present

7th Generation
- Large language models
- Analyst skills augmentation
- Real-time threat mitigation

Future

Resource & Availability Monitoring
- Slow & manual operations
- Sectors: Government, Telecommunications & Military entities

Reactive Monitoring
- Fixed automation
- New sectors: Large Financial institutions & corporations

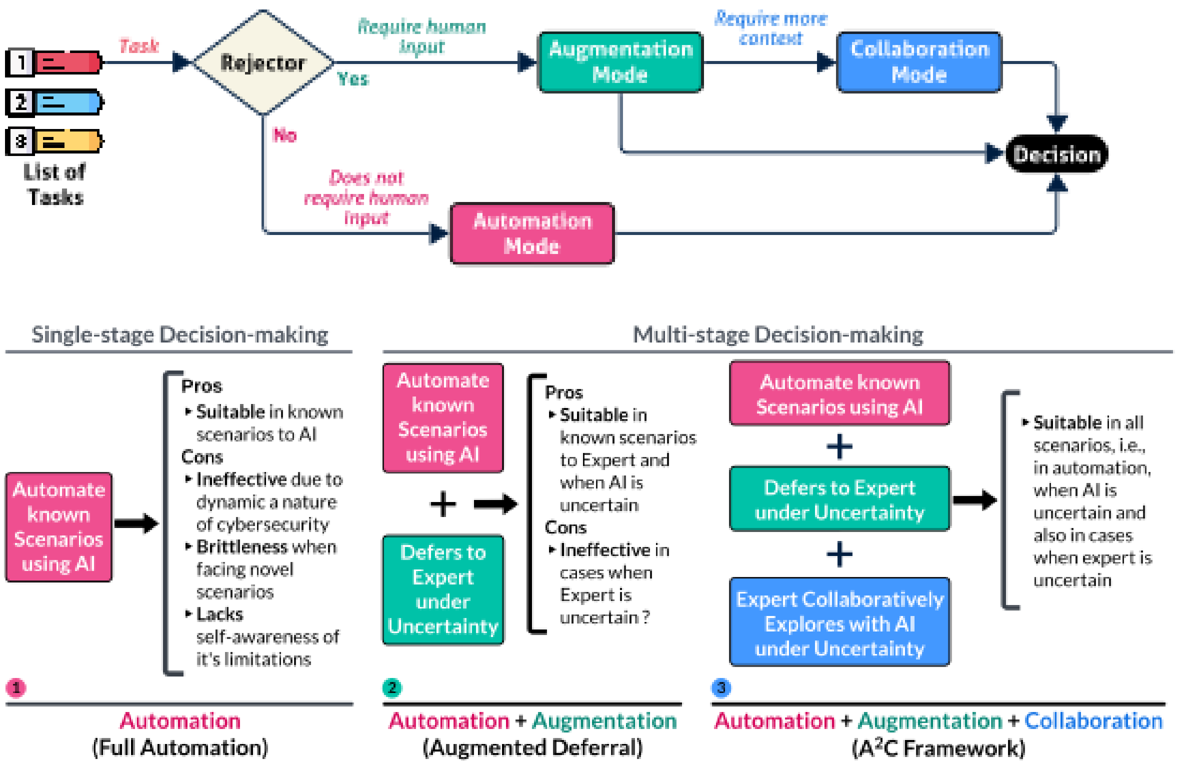
Proactive Monitoring
- Programmable automation
- New sectors: Medium-size Enterprises, Banks & Pharmaceutical industry

Data-driven Monitoring
- Flexible automation
- New sectors: Industry 4.0 & Smart-homes

Collaborative Monitoring
- Hyperautomation
- Collaborative workflows
- New sectors: Healthcare, Industry 5.0 & Metaverse

Shahroz Tariq, Mohan Baruwal Chhetri, Surya Nepal, and Cecile Paris. "Alert fatigue in security operations centres: Research challenges and opportunities." ACM Computing Surveys 57, no. 9 (2025): 1-38.

A²C – A Framework for Adaptive Teaming



1
2
3
List of Tasks
Task
Rejector
Require human input
Yes
Augmentation Mode
Require more context
Collaboration Mode
Decision
No
Does not require human input
Automation Mode

Single-stage Decision-making

Automate known Scenarios using AI

Pros
- Suitable in known scenarios to AI

Cons
- Ineffective due to dynamic a nature of cybersecurity
- Brittleness when facing novel scenarios
- Lacks self-awareness of it's limitations

1

Automation
(Full Automation)

Multi-stage Decision-making

Automate known Scenarios using AI
+
Defers to Expert under Uncertainty

Pros
- Suitable in known scenarios to Expert and when AI is uncertain

Cons
- Ineffective in cases when Expert is uncertain ?

2

Automation + Augmentation
(Augmented Deferral)

Automate known Scenarios using AI
+
Defers to Expert under Uncertainty
+
Expert Collaboratively Explores with AI under Uncertainty

- Suitable in all scenarios, i.e., in automation, when AI is uncertain and also in cases when expert is uncertain

3

Automation + Augmentation + Collaboration
(A²C Framework)

A²C enables dynamic decision-making across three modes:

Automation for routine tasks	**Augmented Deferral** for uncertain cases	**Collaborative Exploration** for complex threats

It adapts to context, uncertainty, and human input.

Modular design allows dynamic switching based on task complexity.

Empirical results show improved accuracy and user satisfaction.

Supports real-time decision-making in phishing and intrusion detection.

Shahroz Tariq, Mohan Baruwal Chhetri, Surya Nepal, and Cecile Paris. "A2C: A modular multi-stage collaborative decision framework for human–AI teams." Expert Systems with Applications 282 (2025): 127318.

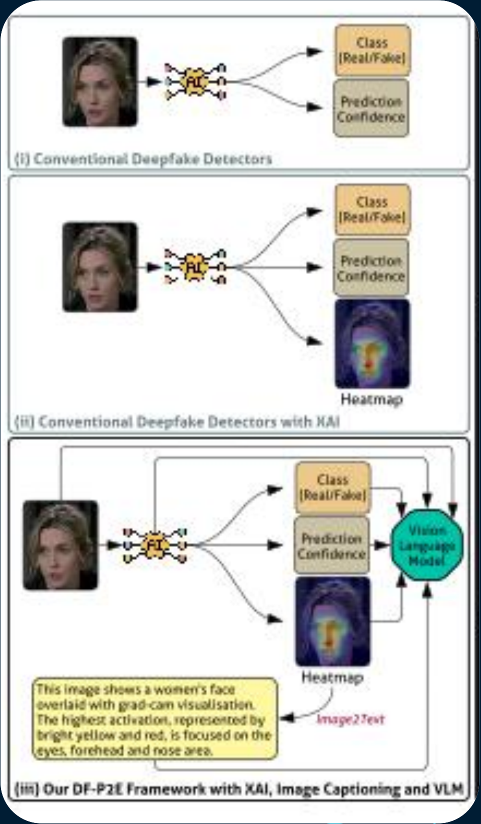
Explaining Deepfakes: A Human-Centered Approach to AI Forensics

Our Framework (DF-P2E) transforms opaque predictions into layered, human-readable explanations.

Visual saliency → Semantic caption → Narrative reasoning

Designed for journalists, investigators, and forensic analysts

Enables validation, questioning, and understanding, not just classification.



(i) Conventional Deepfake Detectors

(ii) Conventional Deepfake Detectors with XAI

This image shows a women's face overlaid with grad-cam visualisation. The highest activation, represented by bright yellow and red, is focused on the eyes, forehead and nose area.

(iii) Our DF-P2E Framework with XAI, Image Captioning and VLM

Shahroz Tariq et al. "From Prediction to Explanation: Multimodal, Explainable, and Interactive Deepfake Detection Framework for Non-Expert Users." accepted for publication in ACM Multimedia 2025 conference

Thank You