Engineering and Physical Sciences Research Council

CRANE

Technology Trends and Horizon-Scanning Report 1

Adam Joinson, Mark Ryan, Emma Walker, Lynne Coventry, and Namrata Primlani

CRANE

Technology Trends and Horizon-Scanning Report 1

25th June 2026

Adam Joinson
Professor of Information Systems
School of Management
University of Bath

Mark Ryan
Professor in Computer Security
School of Computer Science
University of Birmingham

Emma Walker
School of Management
University of Bath

Lynne Coventry
Professor of Cybersecurity
Department of Cybersecurity and Computing
Abertay University

Namrata Primlani
Department of Cybersecurity and Computing
Abertay University

Published by the CRANE Network+.

CRANE is supported by EPSRC Grant Number: EP/Z534845/1

Summary

This report presents the findings from initial foresight activities conducted under the EPSRC Cybersecurity Research and Networking Environment (CRANE) Network Plus between September 2025 and May 2026. The purpose of the exercise was to identify key emerging technologies that are likely to have the biggest impact on cybersecurity over the next decade and priority research areas needed to address this, with a focus on strengthening UK digital resilience.

Using a combination of desk-based research, expert consensus studies and futures workshops, the research focused on 26 technologies across 9 overarching categories that are considered to have the biggest potential cybersecurity impact using a 5-10 year horizon. We examine how these specific technological advancements may intersect with broader systemic factors, including regulatory change, geopolitical dynamics, and human behaviour, to identify key future research priorities in the cybersecurity field.

The findings highlight seven overarching themes and five associated research gaps that will inform future CRANE funding calls. These research gaps include: End-to-end AI trust and verification, adaptive cybersecurity for evolving systems, security of human-centred and human-embedded systems, sovereign and resilient AI infrastructure, and governance and accountability mechanisms.

Overall, these foresight activities establish an initial strategic evidence base for the EPSRC CRANE Network Plus. We use this to identify priority areas for future research, inform policy and funding decisions, and to support the development of coordinated, forward-looking approaches to cybersecurity. Our findings also emphasise the need for proactive, adaptive, and systemic responses to an increasingly complex and uncertain technological landscape and provides a foundation for continued foresight work within CRANE, including the further development and refinement of futures methods specific to the socio-technical cybersecurity context.

Contents

1 Background

As part of the EPSRC Cybersecurity Research and Networking Environment (CRANE) Network Plus, a first phase of technology trends and horizon scanning was conducted from September 2025 to May 2026. The goal of the exercise was to identify future and emerging technologies with a 5-10 year horizon that will have implications for cybersecurity in the following ways:

- The technology or innovation will require new approaches to secure
- The technology or innovation will provide new opportunities for cybersecurity practitioners to improve current approaches
- The technology or innovation could be exploited by adversaries to create novel threats

The approach we took was informed by work on 'futures literacy' (Futureice, 2021; Miller & Sandford, 2018; Miller, 2018), as well as similar exercises undertaken as part of the UK Digital Security by Design programme (e.g., Liveley, 2022). Futures literacy enables "anticipatory activities that are about why and how to 'use-the future' depending on what kind of future we want to imagine" (Miller & Sandford, 2018, p.3). Critically, futures literacy is *not* about predicting what will happen. Instead, it is concerned with *how* we use the future. As Miller (2018) argues, anticipatory assumptions underlie all imaginations of the future and shape present-day decisions in ways that often go unexamined.

The foundations of futures literacy draw on the interdisciplinary study of anticipation, that is, the capacity of systems (biological, social, or computational) to model and act on representations of their own future states (Poli, 2017). Futures literacy argues that most institutions and individuals operate almost entirely in an 'anticipation for the future' mode focused on probable and desirable futures and risk management, while systematically neglecting 'anticipation for emergence' that might allow us to shape possible futures. The result is what UNESCO describes as a "poverty of imagination": a narrowing of the range of futures considered possible, which in turn constrains the choices made in the present (Benedikter, 2022).

Horizon scanning (HS) is a method drawn from futures studies. It is defined as the systematic early detection and assessment of emerging technologies, trends, and threats, primarily for policy makers, across domains including health, agriculture, biosecurity, and environmental management (Cuhls, 2020).

Whereas horizon scanning is generally considered a standalone methodological approach involving the systematic review of various data sources at an early stage, foresight is a broader, more continuous and comprehensive process that incorporates a range of participatory methods and activities (Cuhls, 2020; Miles, Saritas & Solokov, 2016). Horizon

scanning is an *input* to foresight; foresight processes then make sense of it and translate it into strategy.

Cuhls (2020) describes a four-phase integrated model for foresight that guided the technology trends work within CRANE:

1. **Strategic Intelligence Gathering** — searching for information about possible futures, using horizon scanning as the primary method.
2. **Sense-Making** — assessing the relevance and strategic fit of the intelligence gathered.
3. **Selecting Priorities** — using discussion and criteria to identify the most significant directions.
4. **Implementation** — preparing decisions and designing strategies.

Within this model horizon scanning is *necessary but not sufficient* in that it provides the raw material for foresight, but the transformative work of interpreting, debating, and acting on futures signals requires the fuller participatory and analytical processes of foresight, including the capabilities that futures literacy aims to build.

The CRANE Phase 1 Technology Trends and Foresight work used these futures-based approaches to explore future and emerging technologies that may have implications for cybersecurity. Specifically, this work comprised four stages (see Figure 1): (1) Desk research to identify key emerging technologies; (2) Delphi studies to prioritise technologies based on cybersecurity impacts; (3) Futures workshops to identify possible futures, and (4) Synthesis and analysis.

This report summaries the findings from each stage. Fuller documentation is available on request.

2 Stage 1: Desk Research

In the first stage of the process, two researchers with broad-ranging cybersecurity expertise (CRANE co-investigators Joinson and Ryan, with some support from a research assistant) reviewed the following sources for emerging technologies:

2.1 Gartner Hype Cycle for Emerging Technologies 2024 and 2025 and Gartner Hype Cycle for AI 2025

Gartner is a research and advisory organisation that undertakes an annual 'hype cycle' process where emerging technologies are mapped against stages of innovation adoption

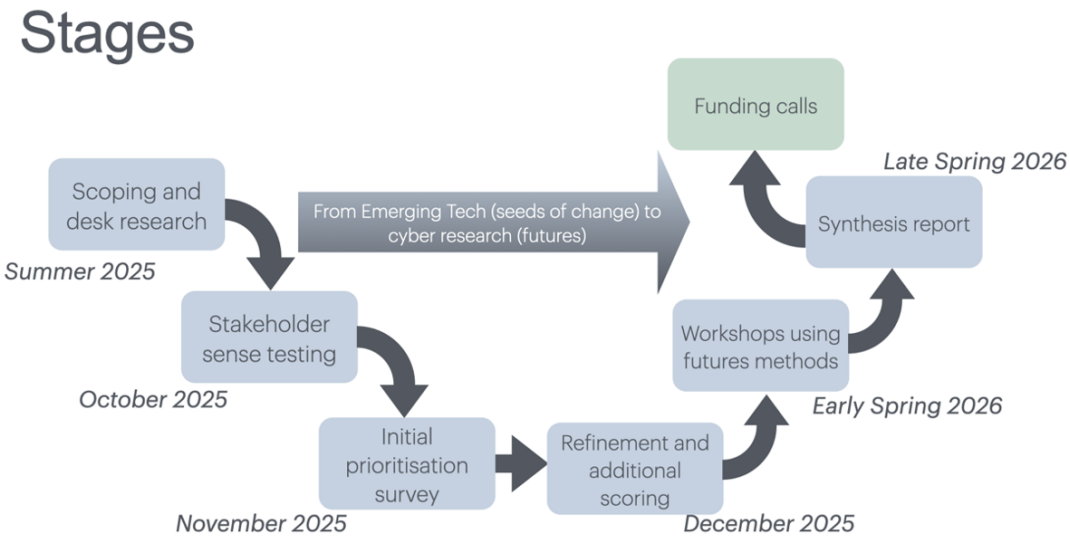


Figure 1: *CRANE Technology Trends stages*

(from innovation trigger through inflated expectations to disillusionment and thence wider adoption).[1] We extracted technologies from the 2024 and 2025 Hype Cycles at the two earliest stages ('innovation trigger' and 'peak of inflated expectations'), as well as the 2025 AI Hype Cycle.

2.2 Deloitte Technology Trends 2020 to 2025

As well as being one of the 'big four' accounting firms, Deloitte provides professional services worldwide and conducts an annual 'tech trends' review.[2] We extracted relevant technologies from the Tech Trends review from 2020 to 2025.

2.3 IEEE Mega Trends 2024 and Technology Predictions 2025

IEEE engages in multiple future technologies processes and outputs, including a Mega Trends "technologies to watch" published in 2024, and an annual "technology predictions" article published in Computer magazine. We extracted relevant technologies from the 2024 Mega Trends article,[3] and the 2025 "technology predictions" article.[4]

[1] https://www.gartner.com/en/research/methodologies/gartner-hype-cycle
[2] https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends.html
[3] https://ieeecs-media.computer.org/media/tech-news/tech-predictions-report-2024.pdf
[4] https://www.computer.org/csdl/magazine/co/2025/01/10834280/23lk5Yd7GkE

2.4 DARPA and ARIA current and emerging topics

We extracted additional relevant technologies from DARPA 'ideas under incubation'[5] and ARIA funding calls.[6]

2.5 Government, academic and learned societies (UK and USA)

UK Government frameworks,[7] Government Office for Science (GO Science) EmTech Team blogs,[8] and reports,[9] alongside searches for relevant areas covered by the Centre for Emerging Technology and Security at the Alan Turing Institute[10], the UKRI Trustworthy and Autonomous Systems Hub[11], the Royal Society[12] and the Royal Academy of Engineering.[13]

2.6 Patents and other sources

Finally, the blog from LexisNexus,[14] including their work on patent analytics was examined, and any gaps added to the list of technologies. We also looked at the programme of work within the Ellison Institute for Technology[15] to identify any significant gaps in our technologies.

3 Selected Emergent and Future Technologies

The product of the desk research was a list of over 80 technologies and innovations. These technologies were scanned by the research team for potential interest to

[5] https://www.darpa.mil/research/ideas
[6] https://aria.org.uk/funding-opportunities/
[7] https://www.gov.uk/government/publications/science-and-technology-framework/science-and-technology-framework
[8] https://foresightprojects.blog.gov.uk/2021/07/29/building-our-vision-for-government-technology-scanning/
[9] https://www.gov.uk/government/publications/engineering-biology-aspirations-report/engineering-biology-aspirations
[10] https://cetas.turing.ac.uk/
[11] https://tas.ac.uk/
[12] https://royalsociety.org/news-resources/publications/
[13] https://raeng.org.uk/
[14] https://www.lexisnexisip.com/resources/ip-blog/
[15] https://eit.org

cybersecurity researchers, as well as discussed with UK Government stakeholders in Emerging Technology, and reduced to 63 for further analysis, with the remainder being either rejected as not having obvious cybersecurity implications, merged with other related technologies, or judged to be too early stage (i.e., beyond the 5-10 year horizon of focus, in which case, they were added to a 'watch list').

For example:

- Compound solar cells, which typically combine perovskite and silicon layers to capture a broader light spectrum with the aim of boosting harvesting efficiency[16] were merged with compound semi-conductors into a single 'compound materials' technology.
- Phage therapy, which involves the use of bacteriophages (viruses that specifically target bacteria) as an alternative to antibiotics[17] was dropped due to no obvious cybersecurity implications.

This final list of 63 was then grouped by the research team with assistance from two AI tools (Claude and Co-pilot, with any discrepancies addressed by the research team), with both a description of the technology and possible timeline added by AI and edited by the research team. The final list and overarching groupings is shown in Table 1 below.

Table 1: *Final technology list with overarching groupings, timeline predictions, and technology description.*

Technology	Description	Timeline Prediction (Mainstream adoption)
Artificial intelligence		
Agentic AI	AI systems capable of autonomous goal-setting and decision-making.	2030-2035
Artificial General Intelligence	AI with human-level cognitive abilities across any task.	2040-2050
Neuro-symbolic AI	Hybrid AI which combines neural learning with symbolic reasoning.	2032-2037
Liquid Neural Networks	Neural networks with dynamic architectures, promising for adaptive AI.	2032-2037
AI-written Software	Software autonomously generated by AI systems.	2030-2035

[16] https://doi.org/10.1038/s41566-025-01732-y
[17] https://royalsociety.org/news/2025/06/summer-science-exhibition-fighting-superbugs/

Technology	Description	Timeline Prediction (Mainstream adoption)
AI-Based Medical Diagnosis	AI tools assisting or automating diagnosis from scans or symptoms.	2032–2037
AI-Guided Pathogen Surveillance	Machine learning systems detecting and predicting disease outbreaks.	2032–2037
AI-Powered Population Health Systems	AI analysing population-scale health data to predict and prevent disease.	2033–2038
AI-Powered Drug Discovery	AI that designs and tests pharmaceuticals faster and more efficiently.	2032–2037
AI for Legal Reasoning	AI systems simulating or assisting legal analysis and decision-making.	2033–2038
AI for Decarbonisation	AI optimising energy, transport, and building systems to support net-zero carbon targets.	2033–2038
Edge AI & Federated Learning	AI models trained locally on devices, enhancing privacy and reducing latency.	2032–2037
Emotion AI / Affective Computing	Systems that detect and respond to human emotions using sensors and algorithms.	2033–2038
Personalised AI Companions	AI entities tailored to an individual's preferences, habits, and emotional needs.	2033–2040
Grief Bots	AI systems designed to simulate deceased loved ones for emotional support.	Unknown
Computing & Data		
Quantum Computing	Computing using quantum bits for exponential speed-ups in certain tasks.	2035–2040
Optical Computing	Computing using light instead of electricity for faster processing.	2035–2045
Neuromorphic Computing	Chips that mimic brain architecture for efficient AI computation.	2037–2045
DNA Data Storage	Encoding digital data into DNA molecules for ultra-dense, long-term storage.	2035–2040
Homomorphic Encryption	Encryption that allows computation on encrypted data without decryption.	2037–2042
Post-Quantum Cryptography	Cryptographic methods resistant to quantum attacks.	2033–2037
Memory-Safe Hardware & Software	Security-focused computer architectures and software preventing memory-related vulnerabilities.	2033–2037
Digital Twins	Virtual replicas of humans or environments that mirror real-world conditions in real time for modeling, simulation, and prediction	2032–2037

Technology	Description	Timeline Prediction (Mainstream adoption)
Spatial Computing	Integration of physical space and digital interaction (e.g., AR/VR, IoT).	2032–2037
Robotics & Automation		
Polyfunctional Robots	Robots capable of switching between multiple complex roles or tasks.	2033–2038
Autonomous Construction Robots	Robots capable of performing construction tasks with minimal or no human input.	2033–2038
Delivery Drones	Autonomous aerial vehicles used for transporting goods.	2030–2035
Fully Autonomous Driving	Vehicles capable of navigating without human input, though deployment is still limited.	2045–2055
Advanced Exoskeletons	Wearable robotic suits to enhance human strength, endurance, or mobility.	2037–2042
Autonomous Space Robotics	Robotic systems capable of autonomously servicing satellites in orbit, including repair, refuelling, and upgrades	2037–2042
Smart Skins for Robots	Flexible sensors and materials that give robots tactile sensing and responsiveness.	2037–2042
Biotechnology & Health		
Engineering Biology	Designing biological systems for new functions, including synthetic life.	2035–2040
Organoid Intelligence (OI)	Using lab-grown brain organoids to perform computational tasks.	Post-2045
Phage Therapy	Use of bacteriophages—viruses targeting specific bacteria—as an alternative to antibiotics.	2037–2042
Nanopore Sequencing	Portable devices for real-time DNA, RNA, and protein analysis, enabling rapid on-site biological analysis.	2032–2037
Point-of-Care Medicines	Medicines produced and supplied directly at the patient's location using rapid, on-site manufacturing technologies.	2037–2042
Programmable Plants	Genetically engineered plants with controllable traits or functions.	2040–2045
Nitrogen-Fixing Cereals	Engineering cereal crops to produce their own nitrogen fertilizer via engineered symbioses or trait transfer.	2042–2050
Bioenergetic Engineering	Manipulating biological energy systems for health, performance, or environmental applications.	2040–2045

Technology	Description	Timeline Prediction (Mainstream adoption)
Neurotechnology & Human Enhancement		
Brain-Computer Interfaces (BCIs)	Direct communication between the brain and external devices.	2035–2040
Neurological Enhancement	Technologies aimed at improving brain function, memory, or cognition.	2040–2050
Targeted Neuroplasticity Training (TNT)	Non-invasive neurotechnology paired with structured training to enhance brain plasticity and cognitive skills.	2037–2042
Extending Human Senses	Technologies that enhance or add new sensory capabilities to humans.	2040–2050
Interspecies Communication using AI	Using AI to interpret and facilitate communication between humans and animals.	2042–2050
Materials & Energy		
Advanced Compound Materials	High-performance materials, including compound semiconductors and perovskite/silicon tandem layers, designed to improve energy conversion and electronic efficiency.	2037–2042
Metamaterials	Engineered materials with exotic properties like invisibility or superlensing.	2040–2045
Self-Healing Materials	Materials that can automatically repair damage without human intervention.	2037–2042
Biodegradable Electronics	Electronic devices designed to naturally break down after use, reducing e-waste.	2040–2045
Spider-Silk Bioelectronic Fibres	Ultra-thin, flexible, biocompatible fibers for wearable health monitoring and environmental sensing.	2040–2045
Small Modular Nuclear Reactors (SMRs)	Compact nuclear reactors designed for safer, scalable energy production.	2035–2040
Energy Harvesting Wearables	Wearable devices generating energy from body movement, heat, or ambient sources.	2035–2040
Persistent Optical Wireless Energy Relay (POWER)	Airborne optical energy relays transmitting energy over long distances.	2042–2050
Environment & Sustainability		
Carbon Capture and Utilisation (CCU)	Technologies that capture CO_2 and convert it into usable products.	2040–2050
Climate Engineering / Geoengineering	Technologies to deliberately alter Earth's climate systems.	Post-2050
Space-Based Solar Power	Collecting solar energy in space and transmitting it to Earth.	Post-2050

Technology	Description	Timeline Prediction (Mainstream adoption)
Smart Agriculture	ICT-based precision farming optimising crop and livestock management for efficiency and sustainability.	2035–2040
Ocean of Things	Networked ocean sensors for environmental monitoring and data collection.	2037–2042
Space & Connectivity		
6G	Next-generation wireless communication promising ultra-fast, low-latency connectivity.	2033–2037
Quantum Sensors & Clocks	Miniaturized quantum devices for ultra-precise timekeeping, navigation, and environmental monitoring.	2037–2042
Smart Dust	Tiny microelectromechanical sensors (MEMS) that detect, record, and transmit environmental data.	2040–2045
Hyperloop	High-speed transportation system using vacuum tubes and magnetic levitation.	Post-2050
Space Tourism Technologies	Technologies enabling commercial travel and stays in space for non-astronauts.	2037–2042
Digital Systems & Society		
Digital Identity & Decentralised ID	Blockchain-based identity systems giving users control over personal data.	2035–2040
Internet-Based Voting	Secure electronic voting for elections and referendums.	2040–2045
Renewed Interest in Cryptocurrency	Cryptocurrency's impact on the stability of the world's financial system.	Ongoing

4 Stage 2: Delphi Studies

A two-stage Delphi process was used, with expert participants primarily from academia, and the public sector being invited to first rank the full list of technologies, and then to provide a deeper analysis of those technologies scored as most relevant to cybersecurity. Delphi is a methodological approach that aims to generate insights from experts in areas with high uncertainty or limited information availability, often using rating scales, rank ordering, or open questions to understand expert consensus as part of an iterative process (Beiderbeck, Frevel, von der Gracht, Schmidt & Schweitzer, 2021).

4.1 First round Delphi

For the first round Delphi, participants were recruited through a request for participation distributed via the CRANE mailing list, RISCS mailing list, alongside specific direct requests to contacts of the research team. The study was also further advertised on LinkedIn (with limited success). Our sampling approach was targeted towards researchers within the cybersecurity field, although we also targeted cybersecurity practitioners within Government and Policy, and received a single response from a cybersecurity R&D lead in industry. The survey was open for two weeks. Participants completed the survey remotely using QuestionPro survey software.[18]

A total of 89 people started the activity. There were 44 completed responses. A total of two people didn't complete any rankings but did volunteer to take part in workshops and provided an email address.

All but one respondent was based in the UK. Occupation was professor (n=15), Reader (n=1), associate professor/senior lecturer/lecturer (n=10), "academic" (n=5), researcher/PhD (n=6), practitioner (n=2), missing (n=3). Gender distribution of responses was 22 reporting as male, 11 as female, 1 as 'other', and 6 selecting "prefer not to say".

The technologies – alongside their description – were placed into groups (as per Table 1), and participants were asked to rank the technologies within each group according to which would have the largest cybersecurity impact in the next 10 years, using the following guidance criteria:

- Will there be cybersecurity research challenges in securing this technology or development?
- Might this technological development create opportunities for adversaries to launch new types of attacks on society or our existing technological infrastructures?
- Does this technological development introduce new security-relevant social (or socio-technical) considerations for cybersecurity research?

If participants were unable to decide on a specific technology, they were instructed to leave the technology in its original position.

Because the technologies were ranked in mixed sized groupings, comparing average ranking across technology groups is not possible since the degrees of freedom to (for instance) rank first or last differs according to the size of the group. As such, both the average rank as well as the number of times the technology was scored in top rank is presented.

To identify technologies for the second-round Delphi, the average ranking in each group, as well as the number of times the technology was ranked first, was considered.

[18] https://www.questionpro.com/us/

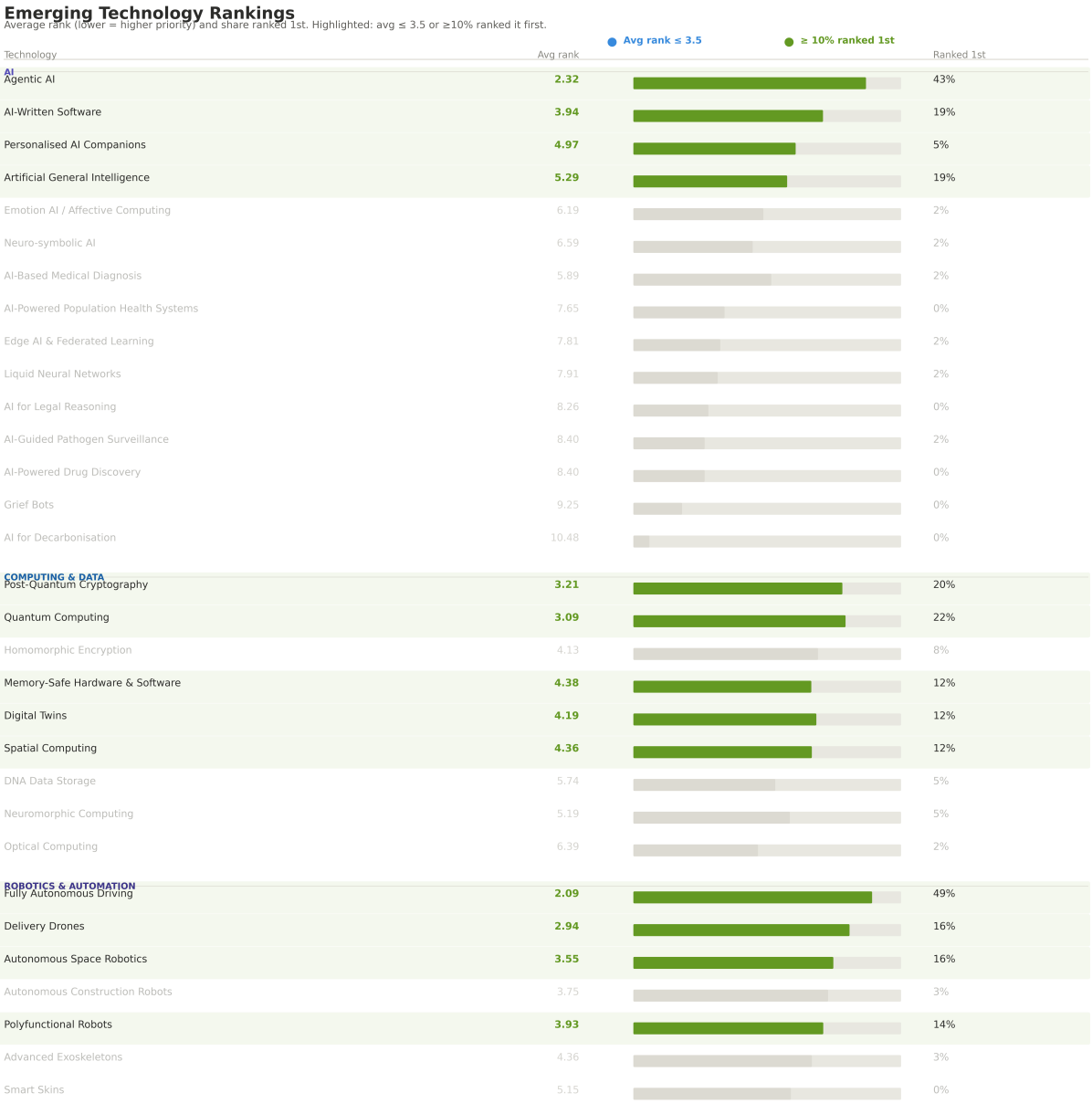


Figure 2: *Ranking (average and proportion of top ranks) per group following first round Delphi (first three groupings of technologies)*

Technologies that were either ranked highly across all participants (higher than 3.5), or those that received a top ranking by at least 10% of participants, were highlighted (see Figures 2 and 3). In the final selection, technologies in groupings with relatively few members were treated more harshly, while those in clusters with many items were treated more leniently, given the *degrees of freedom* issue discussed earlier. This led to 32 technologies being included in the second round (see Appendix 1 for the full list)

4.2 Second round Delphi

For the second round Delphi, the same recruitment procedure was used (i.e. CRANE mailing list), with the addition that contributors from the first round Delphi who had offered to take part in round 2 were emailed directly with an invitation to participate. This led to 26 responses. No demographic information was collected in this round of the Delphi.

Again, the technologies were grouped within the same overarching categories as the first round Delphi, and each technology was scored on the following using a 5-point Likert scale:

- Degree of need for cyber research innovation
- Degree of adversarial exploitation potential

The experts were also asked to rate their degree of certainty in their rankings on a scale of 0-10 after each group of technology. One technology (automated / driverless cars) was missed from the initial survey due to a coding error. As such, ten people were asked to score this technology alone, which was then added to the dataset.

The scores were placed onto a quadrant diagram to identify those that both require cybersecurity research innovation *and* were likely to be exploited by adversaries (see Figure 4).

This initial analysis revealed that – not unexpectedly – artificial intelligence technologies (specifically Agentic AI, AGI, and specific applications of AI such as personalized companions, AI written software, and medical diagnosis) were placed regularly in the upper right quadrant, indicating a high degree of need for research and high adversarial exploitation potential.

There were also several automation technologies (e.g., driverless cars and delivery drones), neurological technologies (e.g. brain-computer interfaces, neurological enhancements), computational developments (e.g. quantum computing and digital twins) and societal use of technology (e.g. internet voting, digital ID) within this upper right quadrant.

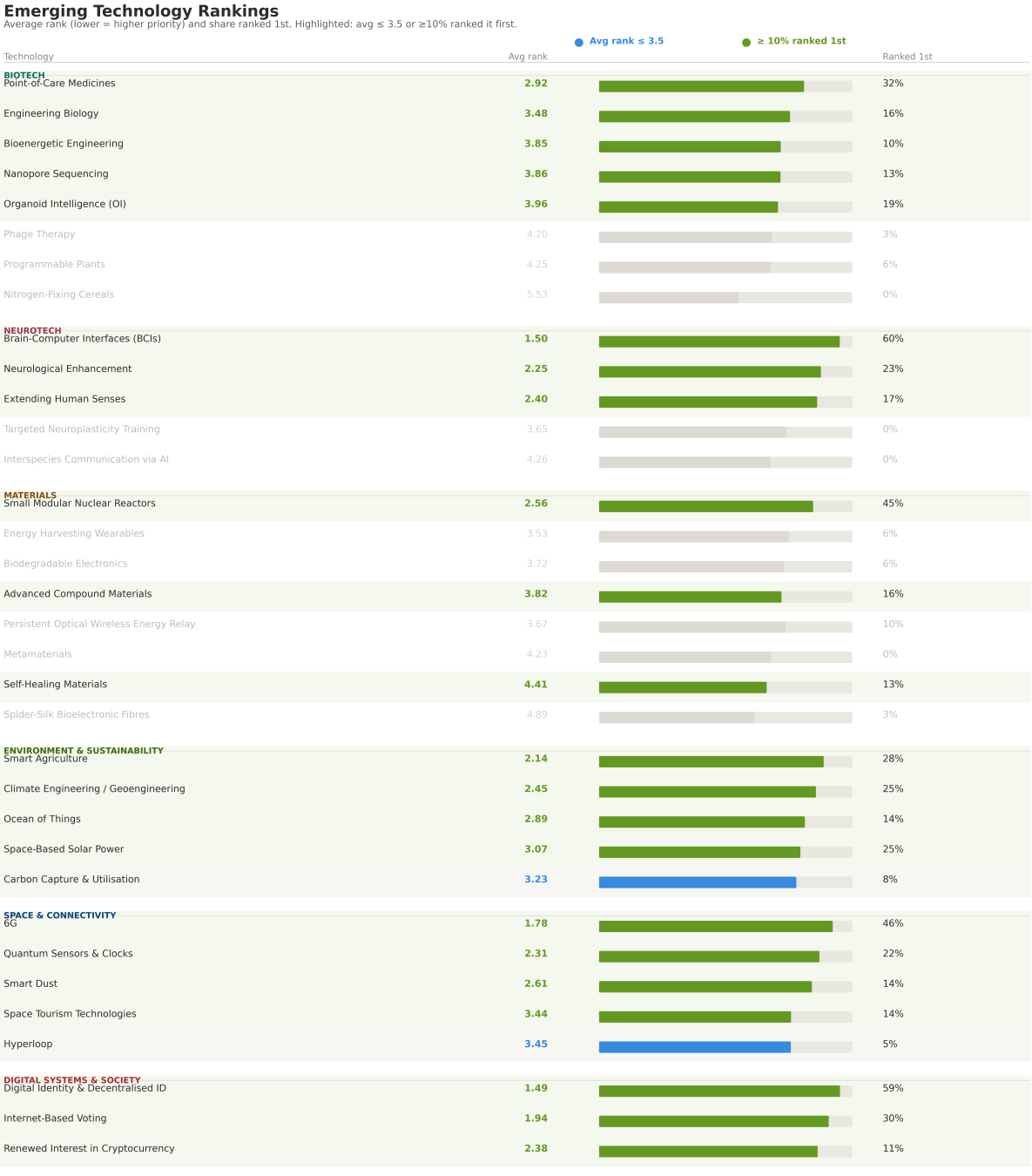


Figure 3: *Ranking (average and proportion of top ranks) per group following first round Delphi (remaining groupings of technologies)*

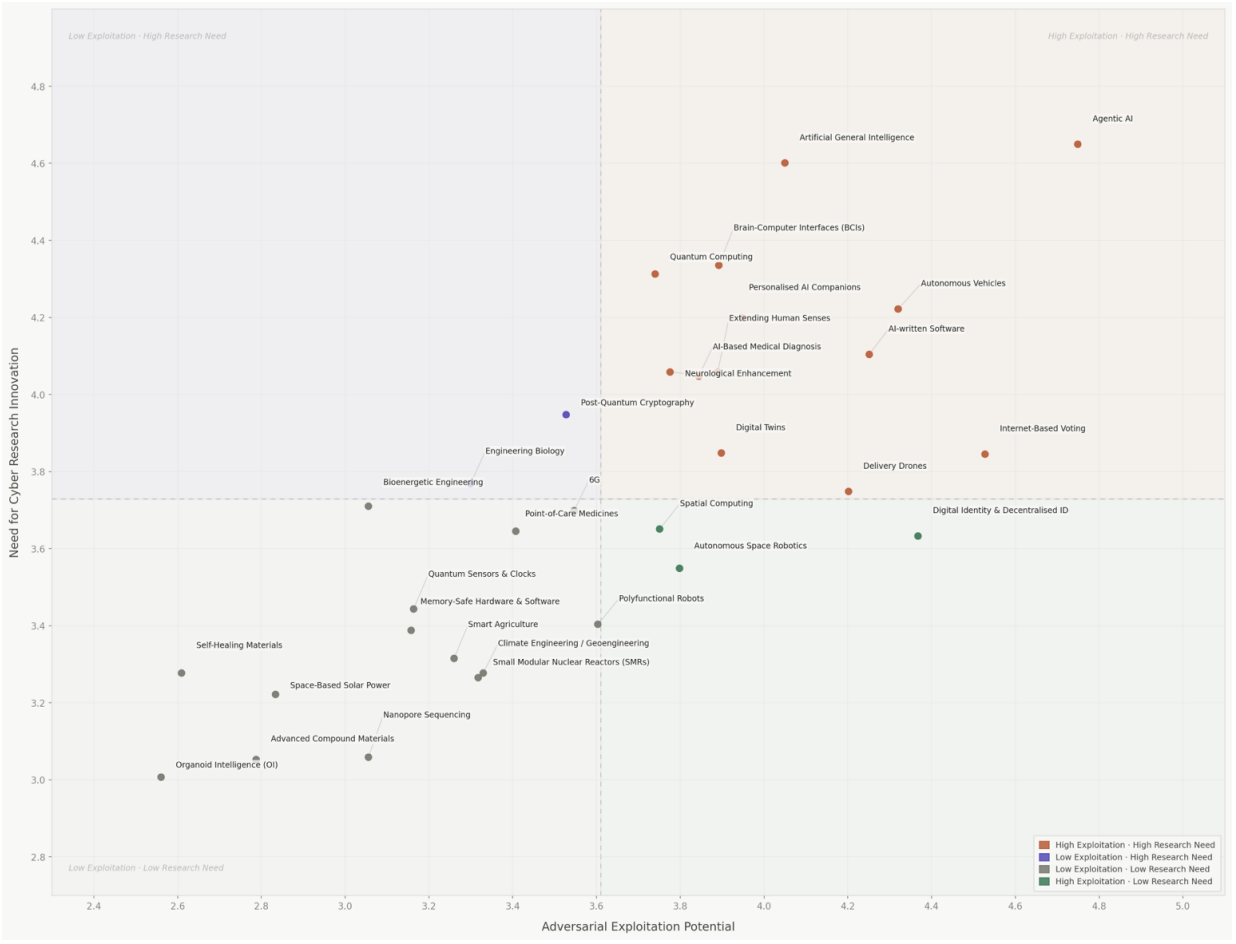
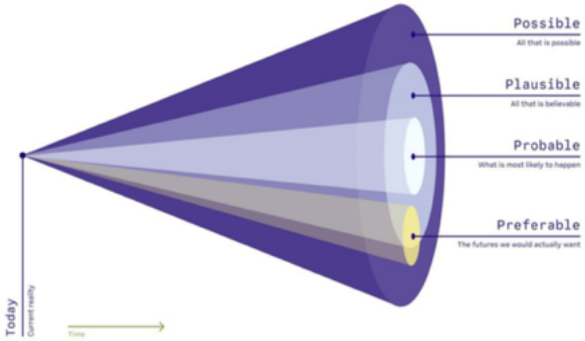


Figure 4: *Quadrant diagram displaying technologies by degree of need for cyber research innovation and adversarial exploitation potential.*



Source: Futurice, 'The Lean Futures Creation Handbook 2.0' https://futurice.com/lean-futures-creation-toolkit

Figure 5: *Diagram of the different elements of futures (e.g., possible, plausible, probable, and preferable).*

5 Stage 3: Workshops

Following the Delphi process, two workshops were run to explore the future implications of the identified technologies in more depth. These workshops were held on consecutive days in London, with invitations sent via the CRANE mailing list, to specific UKRI funded groups focused on the technology in question, and to respondents to Phase 1 of the Delphi who signalled a willingness to take part in future workshops. Two notetakers (both PhD students) also attended the workshops and took notes from the table discussions and group feedback.

Workshop participants were 21 participants in the first (focused on AI) and 19 in the second (focused on robotics and automation). A total of 5 participants took part in both workshops.

The two workshops followed the same format. First, the goals of the workshop were introduced, alongside the relevant results from the Delphi studies. The notion of futures literacy, and the difference between all possible futures, and plausible, probably and preferable futures, was introduced (see Figure 5).

5.1 Opening exercises

Participants were then asked to spend time thinking about two elements:

- "What might the future look like", which focused on considering broader aspects such as global politics, economy, society, environment and technologies in general.

- To develop a “Future headline”, where participants were prompted to sketch a future headline that focused on a potential impact of technology based on the above discussions.

The goal of these two exercises was to encourage participants to begin thinking about a 10-year horizon for technology in general, and to explore both the first and second order impacts of technology (i.e., the more direct and immediate consequences versus the more indirect and delayed consequences).

5.2 'What if' questions and 'implication mapping'

After these two exercises, the workshop used two techniques from the Lean Futures Creation Handbook (Futureice, 2021) to focus on specific technologies discussed within the Delphi studies. The first technique was to work within small groups to explore “what if” questions, where groups combined a chosen technology (e.g., service robots) with either a global or societal challenge (e.g. ageing population).

Within the first workshop, groups were able to choose a technology from the list developed by Futureice, with prompting to focus more on AI or Computing technologies. In the second workshop we limited the range of technologies participants could choose from to those related to robotics and/or automation. During this part of the workshop, participants explored two potential “what if” questions to develop plausible future scenarios. Figure 6 shows a completed “what if” activity sheet.

The second technique involved working within the same small groups to complete “implication mapping”. Firstly, the futures wheel was introduced, which is a method of visualising both first and second order consequences of a particular change, arranged around the PESTLE elements. This provided a structured way for participants to think about potential implications of technology developments in the future.

In small groups, the workshop participants took one of the “what if” statements they developed earlier and used this as the ‘seed’ in the centre of the wheel. They then looked at the potential 1st and 2nd order implications of this statement. Figure 7 shows a completed “implication mapping” activity sheet.

Finally, participants completed a final summation form where they outlined their “what if” question, the associated cybersecurity implications, current tools or knowledge that could be applied to the problem, and potential research areas or gaps. Figure 8 shows a completed summation sheet.

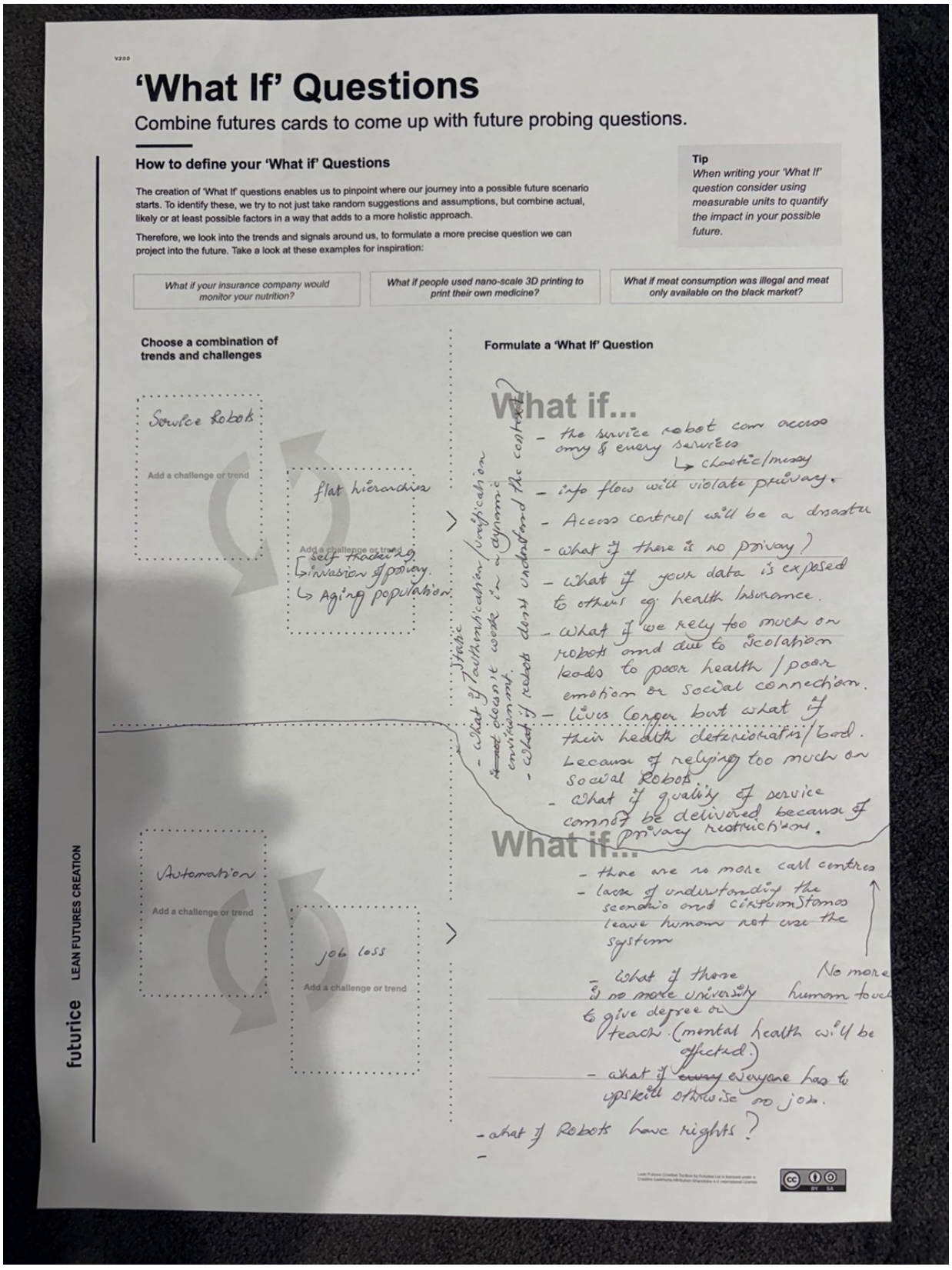


Figure 6: *Example completed "what if" activity sheet*

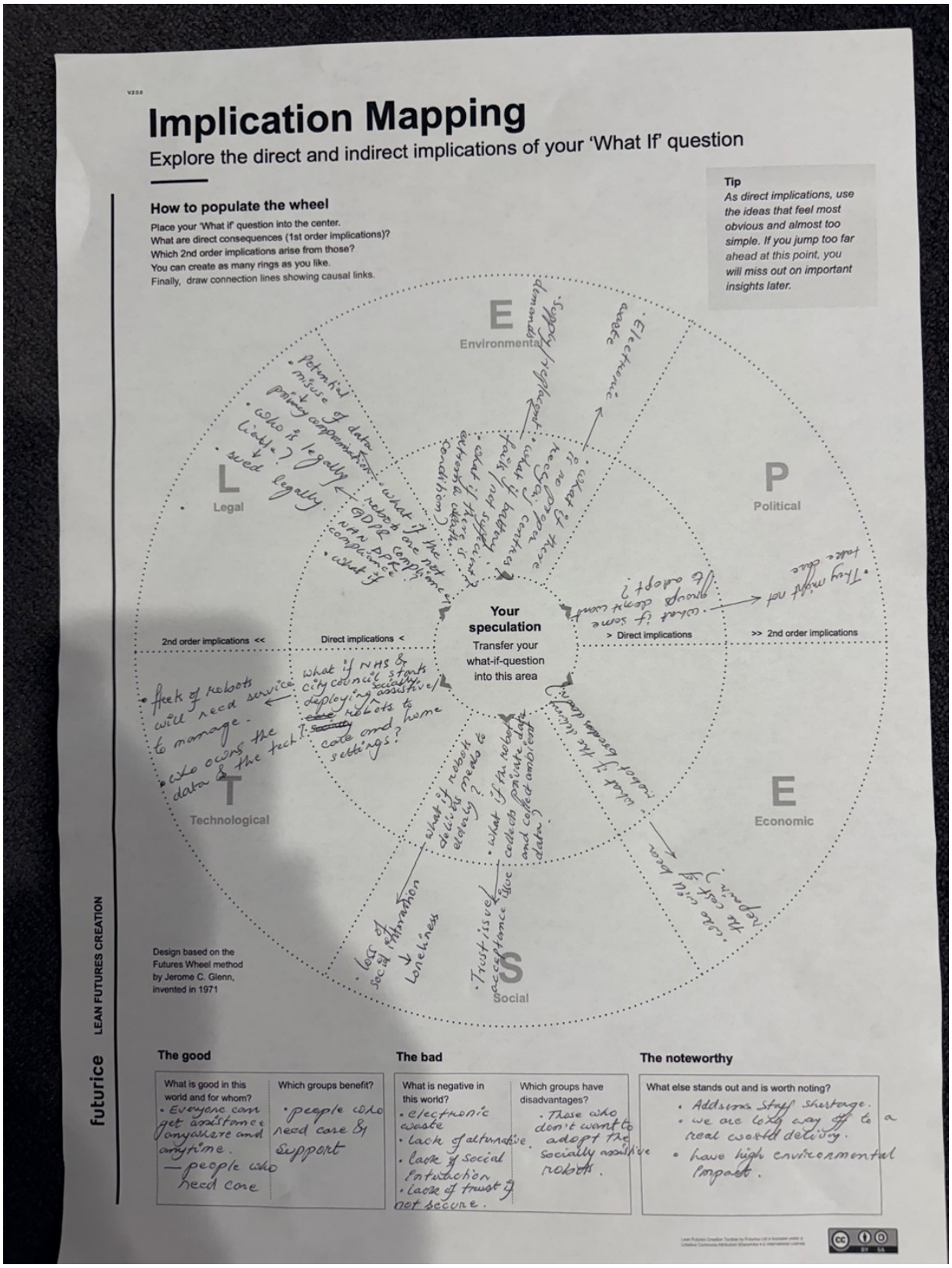
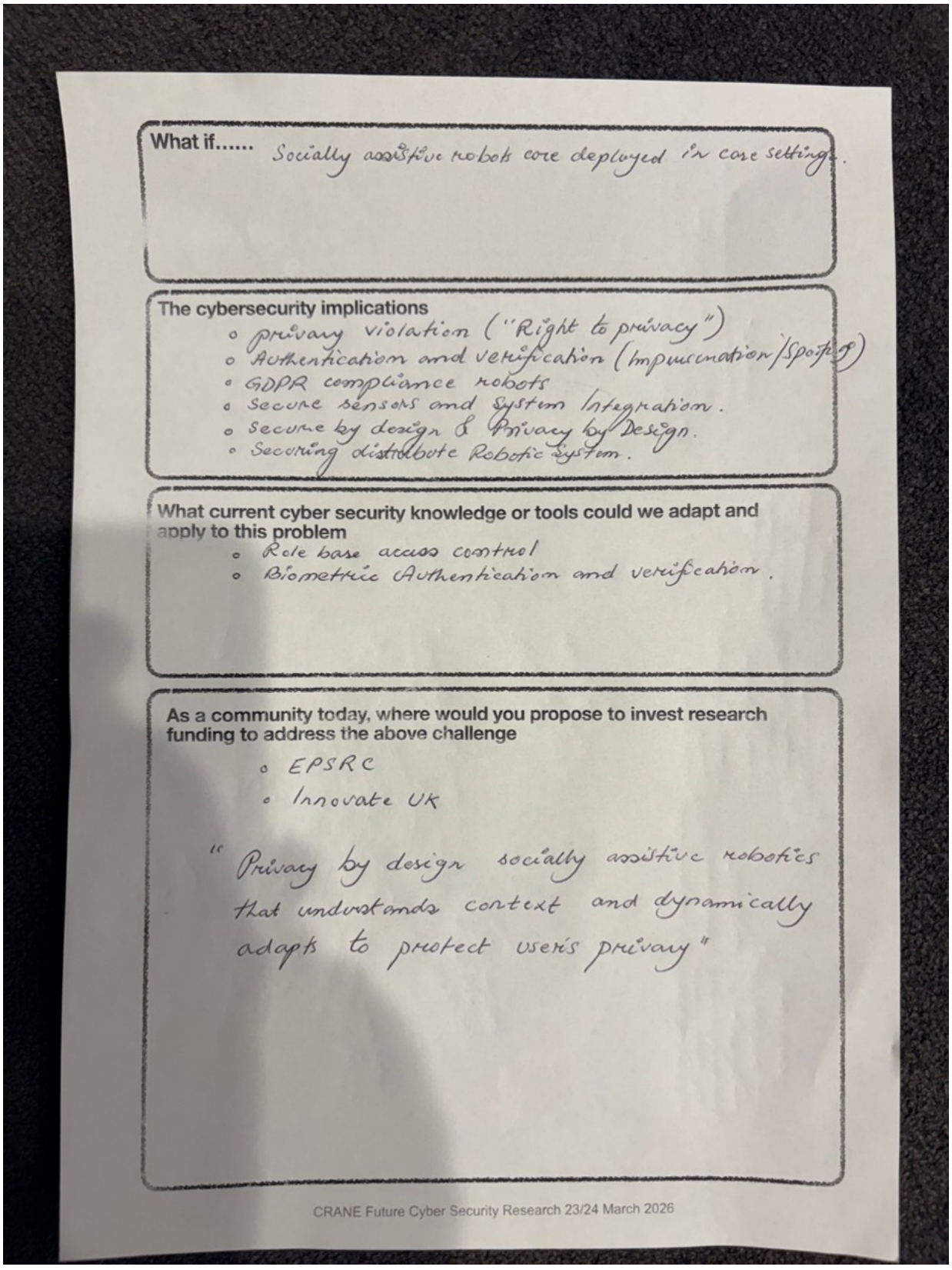


Figure 7: *Example completed "implication mapping" activity sheet*



What if..... Socially assistive robots core deployed in care settings.

The cybersecurity implications

- Privacy violation ("Right to privacy")
- Authentication and verification (Impersonation/Spoofing)
- GDPR compliance robots
- Secure sensors and System Integration.
- Secure by design & Privacy by Design.
- Securing distributed Robotic System.

What current cyber security knowledge or tools could we adapt and apply to this problem

- Role base access control
- Biometric Authentication and verification.

As a community today, where would you propose to invest research funding to address the above challenge

- EPSRC
- Innovate UK

"Privacy by design socially assistive robotics that understands context and dynamically adapts to protect users privacy"

CRANE Future Cyber Security Research 23/24 March 2026

Figure 8: *Example completed summation sheet*

6 Stage 4: Analysis and Synthesis

The notes from the workshops, alongside the activity sheet outputs, were thematically analysed by both an independent researcher and a member of the CRANE leadership team who had attended the workshops as a participant, but had not been involved in the previous stages of the work. The full thematic report is presented in Appendix 2.

6.1 Themes

The thematic analysis initially identified nine themes, which were subsequently synthesised into seven overarching themes.

Theme 1: Post-Trust Information Ecosystems. AI-generated content, misinformation and degraded information quality are eroding traditional notions of trust. Verification becomes harder at scale, while the governance of "truth" and authenticity becomes contested, raising risks for democratic stability, polarisation and radicalisation pathways.

Theme 2: AI Sovereignty, Power and Geopolitical Cybersecurity. Control over AI infrastructure, training data and models is increasingly framed as geopolitical power. Concerns centre on dependency on foreign technology providers and the concentration of control among large corporations, alongside the risk that AI assets (e.g., data centres) become strategic targets in conflict.

Theme 3: Shift to Adaptive, Emergent Cybersecurity Systems. Traditional preventative models are seen as insufficient for non-deterministic, complex systems. Participants emphasised the need for dynamic, learning security approaches able to handle vulnerabilities such as prompt injection, "rogue" or unpredictable agent behaviour, and long-horizon data poisoning.

Theme 4: Cybersecurity Extends into the Human Body (Embodied Cybersecurity). As technology merges with biology (implants, neurotechnology, assistive systems), the attack surface expands into the human body and mind. This creates challenges around bodily privacy, cognitive autonomy, safety-critical failures (including harm/death), and ethical/legal boundaries as human–machine distinctions blur.

Theme 5: Data Ownership, Lifecycle and Death. The analysis highlights the need for lifecycle governance of personal and bio-data—from creation and use, through device obsolescence, to end-of-life and the "afterlife" of data. Issues include post-mortem access rights, enforceable deletion, sustainability and secure disposal/recycling of data-rich devices.

Theme 6: Future Cyber-Physical and Robotic Ecosystems Create Socially Embedded Vulnerabilities. Robots and autonomous systems are mobile, sensor-rich entities embedded in everyday environments, expanding the attack surface across homes,

workplaces, hospitals and roads. Risks scale at swarm and ecosystem levels, and include privacy harms, covert surveillance, systemic failures, and new forms of trust exploitation/social engineering via embodied agents.

Theme 7: Governance, Inequality and Socio-Technical Resilience. Legal and governance systems are widely perceived as lagging behind technological change. The analysis highlights challenges of accountability and liability for AI decisions, cross-border jurisdiction, duty of care, and the risk of widening inequalities (e.g., enhanced vs non-enhanced bodies; vulnerable populations facing increased dependency and reduced agency).

6.2 Cross-cutting research gaps

The thematic analysis identified five clear research gaps.

Gap 1: End-to-end AI trust and verification. Scalable approaches to authenticity/provenance for AI-generated content; clarity on AI "epistemology" (what models know and how to validate it); and metrics for trust erosion, misinformation prevalence and downstream harms.

Gap 2: Adaptive cybersecurity for evolving systems. Formal models and verification methods for non-deterministic, adaptive systems; tooling for self-monitoring/self-healing; and methods to identify emergent vulnerabilities beyond attacker-centric models.

Gap 3: Security of human-centred and human-embedded systems. New paradigms for implants and human-embedded technologies; understanding risks of cognitive manipulation and dependency; and security design that is accessible and usable for diverse and vulnerable users.

Gap 4: Sovereign and resilient AI infrastructure. Frameworks for building and securing "local-first" or modular sovereign AI deployments; end-to-end AI supply-chain security and provenance tracking for models and datasets; and resilience strategies for protecting AI assets and data centres.

Gap 5: Governance and accountability mechanisms. Attribution of harm and liability when AI causes damage; governance frameworks for human-in-the-loop oversight and duty of care; and adaptive, anticipatory governance approaches able to evolve alongside technology.

7 Implications for Future Cybersecurity Research

Taken together, the findings suggest that the most valuable funding opportunities will be those that combine **technical innovation** with **socio-technical and governance**

mechanisms, reflecting the shift to ecosystem security. This supports calls that:

- develop **verifiable information infrastructure** (provenance, authenticity, and integrity at scale),
- build **adaptive security capabilities** for AI-driven systems with emergent behaviour,
- address **embodied and cyber-physical security** where safety, privacy and human autonomy are intertwined,
- strengthen **sovereign/resilient AI infrastructure** including supply-chain assurance, and
- produce **accountability frameworks** that are operational, enforceable, and protective of vulnerable populations.

Following the thematic analysis, a theme ↔ gap matrix was developed using a combination of AI (Claud.ai) and human researcher. Specifically, the Claude was prompted to score the full thematic report using two imagined coders – one of whom focused on "Theme-defining dependency" ('what is structurally required to address the theme's core problem') and one of whom focused on "Systemic leverage" ('what most reduces tail risks and enables credible scaling'). The analysis was cross checked by a human researcher.

Scoring was driven by the following metric:

- **Strong (3)** — Pivotal control point: addressing the gap is **necessary** to address the theme's core challenge **and** the report foregrounds it as central (directly or by clearly equivalent language).
- **Moderate (2)** — Important enabler: the gap is materially helpful, but not the decisive bottleneck/control point for that theme.
- **Emerging (1)** — Plausible but peripheral: the gap is of indirect/second-order relevance; not foregrounded as a major lever in the report.

Both AI coders used the same metric and rated every cell in the matrix as either strong, moderate or emerging. Inter-rater reliability between the two AI coders was analysed and indicated strong agreement: Cohen's κ: 0.953.

Within the matrix, eight theme-gap elements were identified as strong. These were:

- Post-Trust Information Ecosystems ↔ End-to-end AI Trust & Verification
- AI Sovereignty, Power & Geopolitical Cybersecurity ↔ Sovereign & Resilient AI Infrastructure

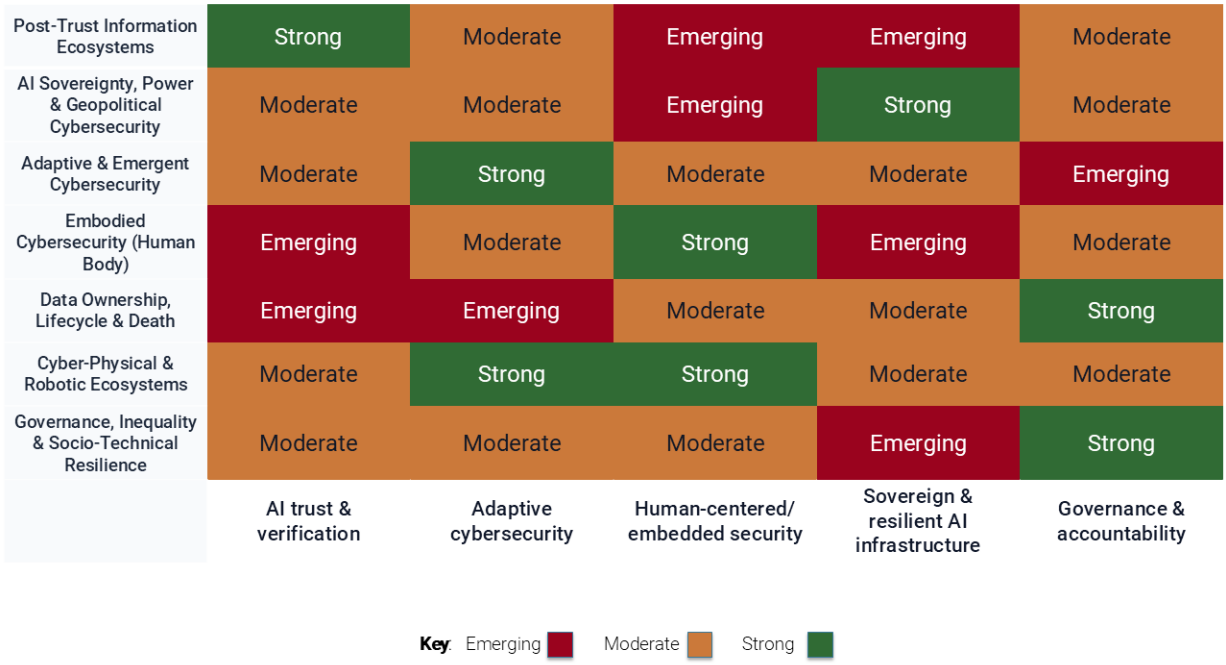


Figure 9: *Theme ↔ gap mapping matrix.*

- Adaptive & Emergent Cybersecurity ↔ Adaptive Cybersecurity for Evolving Systems
- Embodied Cybersecurity (Human Body) ↔ Security of Human-Centred / Human-Embedded Systems
- Data Ownership, Lifecycle & Death ↔ Governance & Accountability Mechanisms
- Cyber-Physical & Robotic Ecosystems ↔ Adaptive Cybersecurity for Evolving Systems
- Cyber-Physical & Robotic Ecosystems ↔ Security of Human-Centred / Human-Embedded Systems
- Governance, Inequality & Socio-Technical Resilience ↔ Governance & Accountability Mechanisms

A detailed consideration of the specific challenges within each of these eight theme-gap areas that contributes to why they are rated as strong, is shown in Appendix 3. Figure 9 displays the theme ↔ gap mapping matrix. This analysis was used as a stimulus material for development of the first funding call from EPSRC CRANE.

8 Conclusion and Plans

This research represents the first iteration of the technology foresight process being undertaken as part of CRANE. It provided an initial testbed for the application of foresight methods to identify specific cybersecurity research priorities that can address key challenges in the coming decade. It also provided a means to engage with a range of stakeholders on this topic and capture generic feedback on the use of these approaches within stakeholder communities and at workshop events – this feedback was overwhelmingly positive from an engagement perspective.

However, the approach of continuously reducing our initial technology list meant that some technologies dropped off the list over time (e.g., opto-electronics) but may prove to have unanticipated near-time cybersecurity impacts. Additionally, the specific expertise and interests of the workshop participants meant that some technologies were not subject to the same degree of deep dive discussion within the workshop context (e.g., bio-engineering). As such, it is important to consider the specific expertise of participants and communicate the potential limitations that this may create in relation to the specific technologies that are prioritised and the specific consequences that are considered.

Since the workshops involved the challenge of thinking 5-10 years ahead, participant discussions tended to focus on broader societal impacts and the challenges associated with this, rather than more specific use cases and outcomes. Further work was needed to translate these into tangible research questions and specific applications and mitigations that may be needed to ensure a cybersecure future for all. However, this is the nature of broad-based futures work and provides a means to prioritise research areas without providing too much constraint from an innovation perspective.

Although AI assistance was used in the analysis of data and creation of the matrix, this work also highlighted that the most valuable insights still came from traditional workshops and survey methodologies. We plan to continue to develop our approach to futures work within CRANE, iteratively developing our understanding of possible, plausible, probable, and preferable futures in relation to the security of emerging technology and influencing the cybersecurity research agenda accordingly.

As part of this work, we will also further develop the methodologies that we use in this area, experimenting with both AI support for futures wheels and the design of more cyber-specific resources. From a human perspective, emerging technologies have the potential to influence behaviour of both malevolent actors and those who they target, in new and evolving ways. As a result, applying foresight methods to consider and mitigate potential behavioural impacts across diverse user groups represents a potential new approach to behavioural research that we also plan to explore.

References

Beiderbeck, D., Frevel, N., von der Gracht, H.A., Schmidt, S.L., & Schweitzer, V.M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX,* 8, 101401.

Benedikter, R. (2022). *In the age of recurring systemic crises, UNESCO Futures Literacy becomes a core global policy task* Global Policy Journal (2022). Retrieved from https://www.globalpolicyjournal.com/blog/06/07/2022/age-recurring-systemic-crises-unesco-futures-literacy-becomes-core-global-policy

Cuhls, K. (2020). Horizon scanning in foresight – why horizon scanning is only a part of the game. *Futures & Foresight Science*, 2(1), e23. https://doi.org/10.1002/ffo2.23

Futureice (2021). Lean futures toolkit. Retrieved from https://www.futurice.com/lean-futures-creation-toolkit.

Liveley, G. (2022). AI futures literacy. *IEEE Technology and Society Magazine,* 14th June, 90-93.

Miles, I., Saritas, O., & Solokov, A. (2016). *Foresight for Science, Technology and Innovation.* Springer.

Miller, R. (Ed.) (2018). *Transforming the future: Anticipation in the 21st century*. Routledge/UNESCO. (Open access.)

Miller, R. & Sandford, R. (2018). Futures literacy: The capacity to diversify conscious human anticipation. In R. Poli (Ed.), *Handbook of Anticipation*. Springer.

Poli, R. (2017). *Introduction to anticipation studies*. Springer.

Appendix A List of Technologies Included in the Second-Round Delphi

A.1 Artificial Intelligence:

1. Agentic AI: AI systems capable of autonomous goal-setting and decision-making.
2. Artificial General Intelligence: AI with human-level cognitive abilities across any task.
3. AI-written Software: Software autonomously generated by AI systems.
4. Personalised AI Companions: AI entities tailored to an individual's preferences, habits, and emotional needs.
5. AI-Based Medical Diagnosis: AI tools assisting or automating diagnosis from scans, symptoms or ambient sensors

A.2 Computing, data and communication

6. Quantum Computing: Computing using quantum bits for exponential speed-ups in certain tasks.
7. Post-Quantum Cryptography: Cryptographic methods resistant to quantum attacks.
8. Memory-Safe Hardware & Software: Security-focused computer architectures and software preventing memory-related vulnerabilities.
9. Digital Twins: Virtual replicas of humans or environments that mirror real-world conditions in real time for modelling, simulation, and prediction
10. Spatial Computing: Integration of physical space and digital interaction (e.g., AR/VR, IoT).
11. 6G: Next-generation wireless communication promising ultra-fast, low-latency connectivity. How likely is it that this technology would be exploited by adversaries in novel ways?
12. Quantum Sensors & Clocks: Miniaturized quantum devices for ultra-precise timekeeping, navigation, and environmental monitoring.

A.3 Robotic and Automation

13. Polyfunctional Robots: Robots capable of switching between multiple complex roles or tasks.
14. Autonomous Space Robotics: Robotic systems capable of autonomously servicing satellites in orbit, including repair, refuelling, and upgrades
15. Delivery Drones: Autonomous aerial vehicles used for transporting goods.
16. Fully Autonomous Driving: Vehicles capable of navigating without human input, though deployment is still limited.

A.4 Biotechnology

17. Engineering Biology: Designing biological systems for new functions, including synthetic life.
18. Organoid Intelligence (OI): Using lab-grown brain organoids to perform computational tasks.
19. Nanopore Sequencing: Portable devices for real-time DNA, RNA, and protein analysis, enabling rapid on-site biological analysis.
20. Point-of-Care Medicines: Medicines produced and supplied directly at the patient's location using rapid, on-site manufacturing technologies.
21. Bioenergetic Engineering: Manipulating biological energy systems for health, performance, or environmental applications.

A.5 Neurotechnology

22. Brain-Computer Interfaces (BCIs): Direct communication between the brain and external devices.
23. Neurological Enhancement: Technologies aimed at improving brain function, memory, or cognition.
24. Extending Human Senses: Technologies that enhance or add new sensory capabilities to humans.

A.6 Materials

25. Advanced Compound Materials: High-performance materials, including compound semiconductors and perovskite/silicon tandem layers, designed to improve energy conversion and electronic efficiency.
26. Self-Healing Materials: Materials that can automatically repair damage without human intervention.
27. Small Modular Nuclear Reactors (SMRs): Compact nuclear reactors designed for safer, scalable energy production.

A.7 Environment and society:

28. Climate Engineering / Geoengineering: Technologies to deliberately alter Earth's climate systems.
29. Space-Based Solar Power: Collecting solar energy in space and transmitting it to Earth.
30. Smart Agriculture: ICT-based precision farming optimising crop and livestock management for efficiency and sustainability.
31. Digital Identity & Decentralised ID: Blockchain-based identity systems giving users control over personal data.
32. Internet-Based Voting: Secure electronic voting for elections and referendums.

Appendix B Full Thematic Report

B.1 Overview

The aim of this workshop was to map research trends in cybersecurity to guide investment and funding opportunities for the CRANE network.

Methods

1. Workshop Structure
    a. 'What if' questions
    b. Implication mapping using PESTLE framework (first and second order consequences)
    c. Designing Newspaper Headlines from the future
    d. Solution Mapping
    e. Research Investment opportunities
2. Thematic Analysis of key themes from transcriptions and workshop outputs

B.2 Themes

Nine themes were identified - Trust, Misinformation and Verification, AI Data Sovereignty, Secure AI, Adaptive Cybersecurity, Bio-hacking, bio-data ownership and death, AI and Aging, Service Robots Security and Privacy, Security for mobile systems.

Further synthesis discussions reduced these to 7:

1. Post-Trust Information Ecosystems
2. AI Sovereignty, power and Geopolitical cybersecurity
3. Shift needed to Adaptive, Emergent Cybersecurity systems
4. Cybersecurity extended into the human body
5. Data ownership, lifecycle and death
6. Future cyber-physical and robotic ecosystems create socially embedded vulnerabilities
7. Governance, Inequality and Socio-Technical Resilience

The discussions highlighted that a key shift in thinking was needed, away from securing systems and networks to securing complex, adaptive, AI-driven ecosystems under uncertainty.

Five cross-cutting research gaps emerged

1. End-to-end AI trust and verification
2. Adaptive cybersecurity for evolving systems
3. Security of human-centred, human-embedded systems
4. Sovereign and resilient AI infrastructure
5. Governance and accountability mechanisms

B.3 Theme descriptions

Post-Trust Information Ecosystems

Core idea: A growing lack of knowing what to "trust", driven by AI-generated content and verification challenges. Poses threats to political stability, public trust and can be a pathway to radicalisation.

Across discussions, trust was portrayed as no longer being a well-defined concept and something that is being structurally eroded and needing to be redefined/redesigned for a future driven by AI-generated content and misinformation. At the same time, verification is becoming technically harder, and less clear who will define "truth" in the future. This theme highlights the need for verifiable information ecosystems that can survive in adversarial conditions.

Topics discussed under this theme include:

Verification challenges: data provenance and authenticity including cryptographic signatures and post quantum verification;

Transparency of knowledge sources: AI model epistemology – understanding what AI "knows" and how AI knowledge can be validated ; and

Misinformation amplification: risk of model collapse – AI trained on AI degrading truth quality

Democratic risks: threats to political stability, public trust, radicalisation and echo chambers.

Research Gaps:

1. A new definition of "trust"
2. Technical/formal verification systems are not yet mature. No robust, widely deployable methods to verify authenticity and reliability of AI generated content at scale. While cryptographic provenance was proposed it may be difficult to standardize globally and may be vulnerable to manipulation of "trusted" sources.
3. Lack of tools to detect long term degradation of AI models over time if trained on "AI slop". May need longitudinal monitoring of model integrity and methods to trace data training lineage and provenance.
4. Lack of metrics to measure misinformation, trust erosion and presence of radicalisation pathways.
5. Need new interventions for content moderation if content being generated by AI.

AI sovereignty, power and geopolitical cybersecurity

Core idea: The redefining of sovereignty, through AI, to data ownership and algorithmic control. Control over AI systems is becoming a geopolitical and societal power issue. Digital sovereignty debated as needed to maintain geopolitical resilience.

Concerns were raised about dependence on **foreign tech infrastructure** and the concentration of power (through ownership and control) in "tech elites" or large corporations. While AI(including autonomous vehicles) was discussed as a **geopolitical asset** and/or a **critical national infrastructure** that could come under attack as a form of electronic warfare.

Concerns were raised about:

- Data sovereignty vs globalised AI systems and dependency on foreign tech providers
- Who has control over Data centres; Training data, underlying models and decision systems
- AI-driven conflict, warfare and national security implications

A number of tensions were highlighted including

- Global collaboration vs national sovereignty;

- Efficiency vs autonomy;
- Innovation vs control

Research gaps

1. No clear framework to build, maintain and secure sovereign AI systems, including a modular approach and Secure "local-first" AI deployments
2. Limited visibility into training data origins and model dependencies could be addressed with end to end AI supply chain security models and provenance tracking for datasets and models.
3. Defence and Resilience strategies for data centres and AI assets (warfare targets) are underdeveloped.

Shift needed to Adaptive, Emergent Cybersecurity systems

Core idea: AI has introduced novel and poorly understood vulnerabilities, many emerging from the technology itself; unintentionally as a result of system complexity and lack of transparency. A new adaptive approach to cybersecurity is required to manage evolving risks from emergent behaviour.

Traditional cybersecurity models were discussed as being insufficient to deal with new threats and an adaptive approach to cybersecurity being required. A **conceptual shift is required** from Preventative security to cybersecurity as a **dynamic, learning system** rather than a fixed architecture.

Discussions around this theme identified that risks are **emergent from the technology** and not always a result of adversarial attacks, as well as constantly evolving, meaning that static "secure by design" will no longer be inadequate

There was a strong emphasis on: **Adaptive cybersecurity** and the need to understand and manage emergent behaviour in adaptive, learning systems.

Critical vulnerabilities identified were

- Prompt injection and LLM-level attacks
- Unknown AI behaviours ("rogue agents")
- Data poisoning over long time horizons

Research Gaps

1. Understanding what security means for non-deterministic AI systems
2. Lack of formal models and verification methods for adaptive security
3. Immature understanding of security for LLMs
4. Understanding of self-healing and self-monitoring AI systems.
5. Ways to identify and understand emergent vulnerabilities (currently over-focused on attackers)

Cybersecurity extends into the human body (embodied cybersecurity)

Core idea: The merging of biology and technology creates deep and interwoven ethical, legal and security dilemmas. As technology extends into the human body, the boundary between human, machine and the environment becomes blurred and could result in loss of bodily privacy, cognitive autonomy and even death.

Synthesising the comments from across both workshops identified the blurring of boundaries between human, machine and the environment. This topic came up in the discussions of biohacking, neurotechnology, AI assisted living and Robotics.

The key concerns raised were the loss of bodily privacy and cognitive autonomy, alongside the expansion of the attack surface into the human body and mind. Security risks exist where devices may malfunction, causing harm or death, and firmware/software vulnerabilities exist within the human body. Critical issues discussed were the Ownership of bio-data (before and after death); the Right to be forgotten in biological systems, and ethical limits of enhancement

Research Gaps

- No established cybersecurity paradigm for implanted devices and future bio-data ecosystems.
- Unexplored risks of cognitive manipulation, behavioural influence and dependency on human-embedded technologies.
- Regulation gaps, including black markets for biohacks, law enforcement and insurance use of bio-data.

Data ownership, lifecycle, and death

Core idea: The ownership and management of data, particularly bio-data, needs a full life cycle approach from creation, to use, to death (of person or device), to the afterlife of data. Concerns were raised about data persistence and the right to deletion, particularly at the end of life.

The key issues discussed were who owns bio-data and device-generated personal data and what happens with such data when a human dies or a device becomes obsolete.

In terms of device lifecycle, concerns were raised about data persistence and the right to deletion, alongside the environmental impact of chips and embedded devices, the need to recycle chips/devices, and how to ensure that no data remains

Research gaps

1. No post-mortem data governance for data after death including ownership of data after death.
2. End of human life data risks
3. Risks of recycling data rich devices.

Future cyber physical and robotic ecosystems create socially embedded vulnerabilities

Core idea: Robots act as mobile, sensor-rich cybersecurity risks which expand the attack surface of daily life while blurring the boundaries of privacy as they are embedded in social environments. Autonomous systems (vehicles, robots, IOT) create systemic, high-impact vulnerabilities. Failure of such systems is not localised, it becomes systemic and potentially catastrophic.

Robotics in new environments such as on the road, in the work office or home introduces physical-digital convergence risks.

Robots/autonomous vehicles etc can act as sensors, actuators and potentially autonomous agents, resulting in mobile, intelligent attack surfaces.

A number of core risks were identified:

- Ubiquitous data collection in homes/workplaces -> privacy invasion;

- Surveillance risks hidden behind “cute” robots or benign devices (trojan horse)
- Swarm-level vulnerabilities (scale amplification); and
- Human–robot interaction risks (trust, manipulation, privacy) and ethical agency (should robots have rights or responsibilities?)

As robots operate in multiple, social contexts (homes, hospitals, workplaces, roads) the inherent technical, social and ethical risks are intrinsically linked.

Research gaps

- Existing cybersecurity not suited to mobile sensing, interacting agents, particularly at a swarm level.
- Data governance for sensor rich environments lacks clear rules regarding collection and ownership of data, access control and real time privacy enforcement.
- Understanding what social engineering and trust exploitation might look like if delivered via a robot.
- How to recognise malicious robot behaviours (eg attempts to manipulate people.

Governance, Inequality and Socio-Technical Resilience

Core idea: Legal and governance systems were consistently described as lagging behind technology. AI driven futures were strongly linked to widening inequalities. The body as part of a cyber-physical system requires governance frameworks that do not exist yet. AI is reshaping human behaviour, relationships and societal norms, the consequences of which are not yet known. AI has having its own motivations and rights is an emerging threats. A clear need for adaptive, anticipatory governance systems was identified.

There were a number of challenges to address under this theme, including

- Where accountability and liability lie for AI decisions and autonomous systems,
- who has jurisdiction over cross border data and bio-data and
- the risk of AI influencing or writing legislation to benefit AI rather than humankind.
- Social consequences including inequality (eg enhanced/non enhanced bodies) and stigmatisation (“outdated bodies”)

Governance gaps are also discussed. In particularly, the lack of frameworks for

- Human-in-the-loop responsibility and control
- The human body as part of a cyber-physical system
- Autonomous decision-making systems
- Non-Ethical and self-preserving AI behaviour

Discussions of the future of AI and robotics predicted widening inequalities. Inequalities in age, wealth, geography and technical capabilities were seen to drive risks including AI driven discrimination, increased dependency on “technology” and loss of personal agency.

Research Gaps

- Attribution of blame when AI causes harm
- Understanding how to frame cybersecurity at a societal level
- How to design security to meet accessibility needs and to protect vulnerable populations
- Governance of human in the loop systems to oversee AI and robots and ensure a duty of care.
- How to safeguard against malicious or exploited humans in the loop.

Appendix C Detailed Challenges Identified within the Theme ↔ Gap Matrix

C.1 Post-Trust Information Ecosystems ↔ End-to-end AI Trust and Verification

The thematic report characterises a systemic transition to **post-trust conditions**, where AI-generated content and misinformation erode the feasibility of relying on informal trust cues and manual verification at scale. The central technical challenge is developing **robust, scalable provenance and authenticity mechanisms** (e.g., cryptographic signing, provenance chains, post-quantum resilient verification) that can operate across heterogeneous platforms and resist adversarial manipulation of "trusted" sources. A second technical challenge concerns **longitudinal integrity**, including detecting model collapse, dataset contamination, and performance drift when training pipelines increasingly incorporate low-quality synthetic content. Socio-technical challenges include defining what "trust" means under contested epistemic conditions, and establishing verification systems that remain legitimate across cultural, political, and institutional contexts without centralising authority over "truth." Policy challenges arise around interoperability and standards (particularly cross-border), while societal challenges involve adoption, usability, and the risk of verification regimes creating exclusion or new inequities (e.g., differential ability to access "verified" channels).

C.2 AI Sovereignty, Power and Geopolitical Cybersecurity ↔ Sovereign and Resilient AI Infrastructure

The thematic report frames AI sovereignty as an increasingly decisive dimension of national and institutional resilience, grounded in **control over data, models, and data centre infrastructure**. Technically, the principal challenge is designing **secure "local-first" or jurisdictionally constrained AI deployments** that maintain capability while supporting auditability, modularity, and secure updates, rather than reproducing opaque dependencies in new forms. A closely coupled technical issue is end-to-end **AI supply-chain security**: provenance tracking for datasets and models, dependency mapping, and assurance mechanisms spanning development through deployment. Socio-technical challenges centre on governance of ownership and control between public and private actors, and the management of trade-offs the report highlights (efficiency vs autonomy; innovation vs control; collaboration vs sovereignty). Policy challenges include procurement constraints, resilience obligations for AI-as-critical-infrastructure, and rules to manage AI assets as potential targets in hostile contexts. Societal challenges include concentration of power

("tech elites"), differential access to sovereign capability, and the legitimacy of who is authorised to control strategically significant AI infrastructure.

C.3 Adaptive and Emergent Cybersecurity ↔ Adaptive Cybersecurity for Evolving Systems

The thematic report explicitly calls for a conceptual transition from static, preventative paradigms to **cybersecurity as a dynamic, learning system** able to manage uncertainty and emergent behaviour in AI-rich environments. Technically, a core challenge is specifying what "security" means for **non-deterministic systems**—including formal models, assurance arguments, and verification approaches that remain valid as models update, drift, or reconfigure. The report's emphasis on prompt injection, "rogue agent" behaviours, and long-horizon poisoning highlights the need for **runtime monitoring, anomaly detection, and adaptive response** that address both adversarial and non-adversarial failure modes. Socio-technical challenges concern the operational acceptability of adaptive controls (e.g., self-healing or automated incident response) and the distribution of responsibility when defensive systems act autonomously. Policy challenges include certification and accountability mechanisms suited to systems whose behaviour changes post-deployment, while societal challenges include trust and legitimacy when security decisions are partially automated and potentially disruptive.

C.4 Embodied Cybersecurity (Human Body) ↔ Security of Human-Centred / Human-Embedded Systems

Embodied systems are a qualitatively different security domain: the **attack surface extends into the body and mind**, raising safety-critical and autonomy-critical risks. Technically, the central challenge is developing cybersecurity paradigms for implants and neurotechnology that integrate **safety, reliability, privacy, and resilience**, including secure firmware/software lifecycles, updates, and failure containment when malfunction can cause physical harm. A further technical challenge is modelling and mitigating **cognitive manipulation and behavioural influence**, which is as yet underexplored but increasingly plausible as human-embedded systems mediate perception and decision-making. Socio-technical challenges include usable security and accessibility: security controls must be comprehensible and operable for diverse users, including vulnerable populations and those in assisted living contexts. Policy challenges arise from regulatory gaps (including black markets for biohacks) and the governance of secondary uses of bio-data (e.g., insurance, law enforcement), while societal challenges include equity, stigma, and contested norms about enhancement and bodily autonomy.

C.5 Data Ownership, Lifecycle and Death ↔ Governance and Accountability Mechanisms

Lifecycle data governance, particularly for bio-data, is a foundational challenge that current regulatory and institutional arrangements do not adequately address. Technically, secure lifecycle management requires methods for **verifiable deletion**, secure decommissioning, and risk-managed recycling of data-rich devices, including scenarios in which devices are obsolete yet retain sensitive data. However, the most binding constraints are socio-technical: accountability across multiple stakeholders (patients, health systems, device manufacturers, cloud hosts) and enforceable rules for post-mortem access and control ("data afterlife"). Policy challenges include creating workable post-mortem governance regimes, harmonising rights across jurisdictions, and balancing deletion rights with public-interest uses (e.g., medical research) under transparent accountability. Societal challenges include culturally variable expectations around privacy after death, family and institutional disputes over access, and the legitimacy of long-term stewardship by public and private actors.

C.6 Cyber-Physical and Robotic Ecosystems ↔ Adaptive Cybersecurity for Evolving Systems

Future cyber-physical systems (robots, autonomous vehicles, IoT) will be **dynamic, networked, and scalable risk surfaces** where compromise can propagate and failures can become systemic rather than local. Technical challenges include adaptive containment strategies (e.g., segmentation/containerisation), resilient swarm-level security, and runtime detection of coordinated attacks or emergent failure patterns in heterogeneous fleets. An increase in socially embedded contexts also suggests increased technical requirements for **real-time privacy enforcement** and secure information-flow control that adapts to setting-specific authorisation and user preferences. Socio-technical challenges include operational governance of updates, incident response, and responsible disclosure in safety-critical environments where downtime and false positives carry substantial costs. Policy challenges concern certification and liability for adaptive systems (including cross-vendor interoperability), while societal challenges concern trust and acceptability when automated security measures influence mobility, care provision, or workplace environments.

C.7 Cyber-Physical and Robotic Ecosystems ↔ Security of Human-Centred / Human-Embedded Systems

Future robots will be socially situated agents: security threats arise not only from code vulnerabilities but from **human–robot interaction**, trust exploitation, and privacy invasion in homes, hospitals, offices, and public space. Technically, this demands security models that incorporate social-context awareness, bystander privacy, and interaction-level threat detection (e.g., recognising manipulation attempts, covert sensing, or "Trojan horse" embodiments). It also implies design requirements for transparent sensing (knowing when recording occurs), and robust access control for sensor data streams that may be repurposed across contexts. Socio-technical challenges include usability and consent in shared spaces; security controls must be legible to non-experts and operational for those who are present but not "owners" of the device. Policy challenges involve regulating surveillance-by-proxy and workplace monitoring, while societal challenges include normalisation of pervasive sensing and heightened risks for vulnerable groups (e.g., older adults receiving care mediated by robots).

C.8 Governance, Inequality and Socio-Technical Resilience ↔ Governance and Accountability Mechanisms

Governance is lagging technological change (a gap that will widen over time), and accountability is central to resilience under AI-driven transformations. Technically, accountability increasingly depends on auditability (traceable decision pipelines), assurance cases for autonomous decision-making, and mechanisms that constrain or verify behaviour in human-in-the-loop systems. Socio-technical challenges are dominant: attribution of blame when AI causes harm, assignment of responsibility across developers/deployers/operators, and designing security controls that are accessible and protective of vulnerable populations rather than deepening inequality. Policy challenges include enforceable duty-of-care regimes, cross-border jurisdiction for data and bio-data, and safeguards against regulatory capture or misaligned incentives where governance frameworks are shaped by those who benefit from weaker accountability. Societal challenges include legitimacy and trust in institutions, as well as distributional impacts (e.g., enhanced vs non-enhanced bodies, owners vs users), which may amplify dependency and reduce agency if governance is not anticipatory and adaptive.