

Security and Ethics of Human Augmentation workshop

Summary Report - March 2022



Contents

3
3
4
5-7
8
8-11
12-15
16-19
19-21
22-26



Introduction

Human augmentation technologies (HATs) raise significant ethical issues because they affect our fundamental understanding of what it means to be human. With their potential to connect humans digitally to the Internet of Things, they also raise significant security issues.

As with any technological advance, the speed of discovery can outpace society's capacity to handle the consequences of its application. Malicious negative uses are, unfortunately, as likely as pro-social positive ones. What's more, there are plenty of opportunities for unintended systemic impacts both good and bad.

With HATs of various types growing in maturity and attracting increasing amounts of R&D investment, now is the time to consider the possible effects in an attempt to regulate them for the public good and to protect our national interest.

Organized by KTN's Neurotechnology Innovation Network in collaboration with national security and defence partners, the Security and Ethics of Human Augmentation workshop held on 18 November 2021 brought together a group of experts in the field of various HATs to consider the medium-term impacts in a range of scenarios.

Objectives

The objectives were to:

- Share our collective understanding of the scientific and technical landscape
- Gather intelligence to inform senior decision-makers and policy-makers
- Inform the strategic direction of future research in industry and academia
- Identify the best potential areas for future action



Structure

The half-day event was structured around a series of three wholegroup presentations, each followed by break-out sub-group discussions.

The whole-group presentations set the scene for the discussions in the sub-groups.

The three questions were:

- 1. What opportunities and threats could HATs present for personal and national security?
- 2. What technology areas need to be protected in the UK from a prosperity and security perspective?
- 3. How do we ensure HATs have a positive impact on society, and what are the key ethical principles that will ensure HATs are used equitably and safely?

The event operated under the Chatham House Rule to encourage participants to speak freely.



Background

The context for the workshop is given by three recent publications:

- 1. Human augmentation: dawn of a new paradigm: a strategic implications report from the Ministry of Defence
- 2. Human Performance Optimization and Enhancement from the Multinational Capability Development Campaign (MCDC)
- 3. National Security & Investment Act (NS&I) 2021, which establishes a new, stand-alone statutory regime for UK government scrutiny of, and intervention in, acquisitions and investments for the purposes of protecting national security

HATs conceive of the human body as a platform for augmenting physical, psychological and social performance, i.e., enhancements above humans' natural biological potential. There are several modes of operation, including biological, chemical and cybernetic.

HATs have the potential to infringe individual rights, leading to degradation, exploitation, control and limitation – ethical considerations that were emphasized during discussions.

HATs blend our understanding of human physiology, biochemistry and psychology with other areas of science and technology.

Taken from the reports listed at 1 and 2 above, the topics include, in alphabetical order:

- Al and system security
- Bio-informatics
- Biomonitoring and sensing
- Capability development
- Data collection and analysis
- Genetics germ-line (heritable) and somatic modification
- Gut microbiome and nutrition
- Human-machine teaming with neural interfaces, including invasive and noninvasive brain interfaces
- Language translation

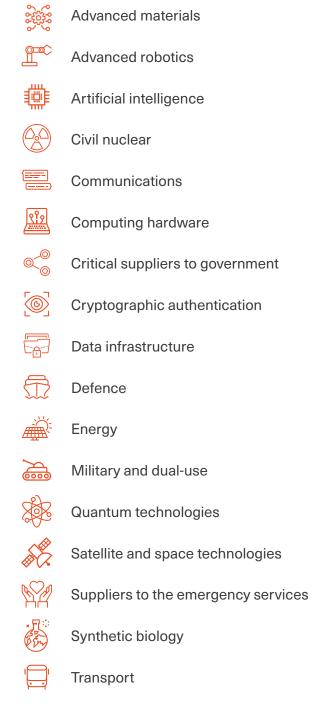
- Neurotechnology
- Optimisation methods
- Passive and powered exoskeletons
- Pharmaceuticals
- Sensory enhancement and communication, including with implantable biosensors and augmented/virtual reality
- Smart textiles
- Tele-existence
- Training

0

1 0 0 11 0

01000

The NS&I identifies 17 areas of technological development that hold particular national security sensitivities. They are, in alphabetical order:



As can be seen, there is considerable overlap between the NS&I list and the list of science and technologies that underpin HAT, which is a measure of the potential for HATs to entail security concerns. 11 1 0 1 1 0 1

Key issues

With the potential for two-way digital communication and control between humans and machines/computers that depend on a complex, multi-part, multi-ownership digital infrastructure, HATs raise serious issues, including:

- Security, including 'cyber-biosecurity': How to prevent hacking, including by those intent on committing crimes, seeking political advantage, or damaging national interests?
- 2. 'Cyberworld' two-way flow of data and thus influence: How to protect human rights to autonomy and agency?
- 3. Data ownership, privacy, use: How to protect against breaches of privacy and unfair or malicious exploitation of data, including to commit crimes, exert political influence, or undermine national interests?
- 4. Lack of experience and effective regulation: How to assure against technical failures and unintended consequences?
- 5. Legacy risk: How to ensure continuing functioning of tech that humans depend on for their quality of life, especially of implanted technology?

The underlying principles of technology development

Given that technology is ethically neutral (with the potential to be used for good or bad) and cannot be un-invented, the following principles are thought to be important for mitigating negative ethical and security impacts:

- Ethical guidelines for technological development, policy and regulation should be established and applied now to mitigate problems later
- Ethical considerations are best protected through transparent co-design with security experts, technical specialist, clinicians, patient groups, end users and others.
- Security by design should be the default starting position for any technology developer

100

01

Break-out session 1

Possible scenarios in 2030

Participants in the first (of a total of three) break-out session were charged with examining three scenarios in 2030 for the opportunities and threats HATs could present for personal and national security. Specifically, they were asked to identify the good and bad impacts.

Scenario A Impact of neurotech on the NHS

Scenario description:

- Simplistic but effective brain-computer interfaces (BCIs) are available on the NHS
- Symptoms of conditions like Parkinson's Disease and depression reduced as standard
- Despite warnings from experts, there are some examples of neurotechnology used to improve cognitive abilities of patients in private clinics, dubbed by some as 'cosmetic surgery for the mind'
- Unregulated devices are being sold under the banner of "wellbeing", "fitness" and "cognitive boosters"
- Neurotech healthcare tourism available to China, where invasive BCIs are in wider use, but with mixed outcomes
- A black market of neurotech is growing, with sometimes horrific consequences for individuals
- Recent hacking of a neurotech healthcare company revealed vast amounts of neuro telemetry
- Hospital trials of neural dust in critically ill patients is opening up revolutionary new treatments



Commentary on scenario

There were a variety of views about whether scenario is realistic.

- Has good potential for considerable health outcomes, but many developments will take longer than 2030
- Invasive technologies such as deep brain stimulators are already providing positive health outcomes with far more developments likely by 2030 (e.g., bioelectronic medicines)
- Reading thoughts via BCIs will not be mainstream by 2030The complexity of neural signals in the brain means that this sort of technology is unlikely to emerge even by 2050
- The focus will be on disease modification/treatments, not augmentation, except maybe non-invasive technologies

Positive impacts

Technical and medical factors:

- Don't have to offer free entries
- Has good potential for considerable health outcomes by offering treatments for many chronic conditions currently beyond pharmaceutical options: e.g.
- Combining AI and neurotechnology could boost cognitive ability which could have medical and wider societal benefits
- NHS can integrate it as monitoring tech to check on patients post-discharge from hospital/treatment

Structural factors:

- Applications in NHS allows access to be inclusive
- Rigour of testing and regulation ensure that applications in NHS (and its infrastructure) are effective/safe
- ✓ Adoption of HATs in NHS helps to build trust in public
- Neurotech already approved for medical uses is easier/safer to adopt in non-medical uses
- Regulatory pathway in the UK, especially for ethical concerns, is one of the best in the world



Negative impacts

Data security factors:

- × Cybersecurity is important for neuro-digital applications but hacking vulnerability can lead to serious harm
- X Devices that connect to BCIs and the wider infrastructure are open to hacking (e.g., DDos attacks) – swamping NHS systems will damage health service and prevent BCI devices from being put to intended use
- × Systems (e.g., Bluetooth) open to hacking to affect HA patient
- × Who owns data patient, company, NHS?
- × Data owned by commercial concerns
- \times Where is data stored?
- imes Who can access ('decode') the data?

Technical and medical factors:

- × Unnecessary treatment. Example: when depression is healthy response to external stimuli, treating patient (rather than addressing external stimuli) is potentially unethical
- V Unstable, 'noisy' IT infrastructure affected by interference from external sources (e.g., pacemakers affected by microwaves/induction hobs, or sounds from other devices picked up in implanted hearing aid)
- Extreme challenge of validating/verifying neurological data
 can only be done for few, simple data
- X How to match intervention to the patient is not fully understood
- Unknown risks long-term effects, compatibility with other factors, etc.
- × Relative benefits unknown, especially in comparison with rival (e.g., pharmacological) treatments



Social factors:

- × Risk of social inequality: since not everyone can or would want to be augmented/enhanced, some humans will be left behind
- × End-user approval
- × Societal approval
- × Risks associated with medical tourism ground rules vary in different territories/jurisdictions
- X Autonomy and agency e.g., if mood can be artificially modified, who gets to say whether target moods are good or bad?
- × Invasion of privacy if thoughts can be read

Structural factors:

- X Development of neurotech is continually chasing funding, which distracts from discovery research. Longer term UK research programmes could provide more funding stability for academia, industry and clinicians
- × Regulatory pathway can be very challenging and slow, which lets other countries overtake, creating commercial and security risks, or tempting developers to bypass it and target other markets, creating safety and ethical risks. There are regulatory opportunities with this emerging technology area which are currently being explored by the Regulatory Horizons Council
- imes NHS processes mean that adoption of approved tech is slow
- Approved neurotech is not easily adopted by clinicians lack of IT infrastructure to protect against disclosure and identity theft



Scenario B Blurring reality – the metaverse is a reality

Scenario description:

- AR/VR with haptic access to the metaverse has passed early adopters phases several years before and is now being used by the masses
- Professional gamers driving non-invasive BCI market for the edge on reaction timings in global tournaments with prizes over \$100 million;
 - Use of invasive BCIs banned due to health concerns
 - 'BCI doping' persists with biodegradable neuro-dust difficult to detect in testing
- AR/VR allowing many to connect and work like never before as the norm
- Users spending big money in the metaverse
- Poorer sections of the global population being left behind due to less immersive access
- Several cases of exploitation in and for the metaverse
- There is rising tension between norms in the physical and immersive virtual worlds, including governance and societal structures
- Conflicting studies are emerging on the mental and physical health impact of the metaverse



Positive impacts

Security factors:

- Quicker responses to public threats (e.g., terrorist attacks/ serious crime)
- More data makes it easier to spot bad actors
- Improved emergency responses virtual mission rehearsal allows better coordination, situational awareness
- ✓ Gaming military operations could improve military outcomes
- ✓ Removes humans from frontline, protecting them from harm

Legal factors:

Legal liability easier to establish ('dashcam in pocket')

Social factors:

- Better employment access and opportunities for people with psychological or physiological disabilities
- Improved data analysis, threat response, training
- Remote surgical robots improve access to healthcare

Commercial factors:

- More efficient testing and quicker product development test on virtual models before committing to physical ones
- Growth of market for gaming



Negative impacts

Security factors:

- × Individuals' experience of metaverse can be spoofed
- × Unregulated body hacking (through BCIs)
- Potential to make people do things they did not intend for malicious ends
- imes Security threat of identity fraud
- × Vulnerability to malicious (signal jamming technology) or accidental interference (failures)

Legal and structural factors:

- imes Identity harder to validate
- X Legal difficulty can there be real-world liabilities for actions in the metaverse?
- × Reach regulatory consensus and making regulation enforceable is a huge challenge

Health factors:

- × Extreme addiction
- × Technology jeopardizes health and wellbeing if not classed as medical devices and is subject to different, lesser regulation
- × Gaming military operations could affect mental health
- Unintended side-effects (e.g., neural dust builds up and causes toxic shock)

Political factors:

- × Easier to target individuals with propaganda
- imes Propaganda easy to produce and promote



Social factors:

- Social inequality different ability to access metaverse (i.e., devices and necessary telecoms infrastructure) will lead social inequalities between classes and/or nations; could become deeply entrenched if used in school education
- Cultural inequality cultural differences could affect engagement, building up inequalities
- × Moral standards drop because society is used to and inured against lesser consequences of bad behaviour in metaverse
- × Encourages 'catfishing' i.e., luring someone into a relationship by means of a fictional online persona
- imes Isolation from real-world social interaction
- imes Dissociation from real world impacts social cohesion
- × Reduction in social skills and physical wellbeing, leading to higher healthcare costs



Scenario C Augmented soldier

Scenario description:

- The first avowed, large-scale combat deployment of augmented soldiers with exoskeletons and other augmentations
- AR allows drone and rear vision for the first time;
 - through simplistic cybernetic eye for some
- Adversaries considering asymmetric warfare approaches whilst readying their own augmented soldiers
- Wounded soldiers returning to combat with augmentations they wouldn't have accepted prior;
 - Different militaries viewing ownership of the augmentation differently
- Some soldiers opting out of augmentation
- Mental wellbeing effects being uncovered and researched
- Augmentation of non-frontline soldiers commonplace for human-machine teaming, EDI and other purposes;
 - Sometimes surpassing known human capabilities
 - Other groups feeling pressured to take augmentation to keep up and some feeling left behind

This scenario was considered by two break-out groups. One chose to consider front-line personnel, the other non-front-line personnel.

Positive impacts for front-line military personnel:

- Makes it easier to recruit soldiers enable people who might otherwise be excluded to join
- Improves soldiers' capabilities
- Can detect soldiers' emotional problems, fatigue, mental capacity (burn-out)
- Can monitor performance and medical status remotely
- Can capture bad behaviour
- Soldiers are easier to track



Negative impacts for front-line military personnel:

Human rights factors:

- \times Threatens soldiers' rights in various ways:
 - Will they be able to refuse augmentation?
 - Will they have personal autonomy/free will?
 - To what extent can commander override autonomy ('human robot'), police thoughts ('detect extremist views')?
 - To what extent is privacy breached through remote monitoring?

Duty of care factors:

- Requires careful definition and application of duty of care because of:
 - young age
 - potential for soldier to be disenfranchised
 - potential for soldier to be less risk-averse, putting them in more danger ('cannon fodder') than is warranted

Health factors:

Potential for physical or emotional harm from leaving HA in body after it is needed or from removing it

Social factors:

- X Is there jeopardy in augmented ex-service people being at large in society?
- Public disapproval, especially if augmented soldiers' human rights are seen to be infringed

Political and security factors:

- Loss of sovereign capability if the UK falls behind developments by foreign actors
- X Threats to security because of errors in tech and supporting infrastructure
- X Tech that allows soldiers to be tracked and convey intelligence can be hacked by enemy actors



Positive impacts for non-front-line military personnel

- HAT can help to match individuals to roles based on their abilities (e.g., cognitive abilities)
- Knowledge can be 'uploaded' to humans affects training/ education

Negative impacts for non-front-line military personnel

Security factors:

- X The more 'avenues for data' (e.g., tracking individuals) there are, the more vulnerability: both the augmented individuals and the supporting technological system and data flows (telecommunications, supply chain, etc) can be targeted (e.g., EMP electromagnetic pulse 'turning off' augmentations)
- X Commercial interests lead, not government. Their interests are not aligned. Al and algorithms (e.g., for targeted advertising, propaganda) already at high TRL but not controlled by governments

Social factors:

- Unfair pressure (from organization or peers) on personnel (e.g., analysts) to enhance cognitive capability through HA. (This already happens in other work spheres – cocaine use to keep going in some high-pressure work environments)
- × If not reversible, HAT implants can close off options for individuals, who might later have regrets
- X HAT companies with data about HAT users know more than is fair or safe about the users

Regulatory factors:

- × Regulation does not capture all HATs
- Little regulation about use of data derived from augmented humans – government regulators can't keep up and, because there is no public debate, the public accepts it by default. (It is noted that the risk is not certain: public backlash halted the adoption of GM technology)
- Regulation can easily be ignored if not policed and if sanctions for breaches don't deter

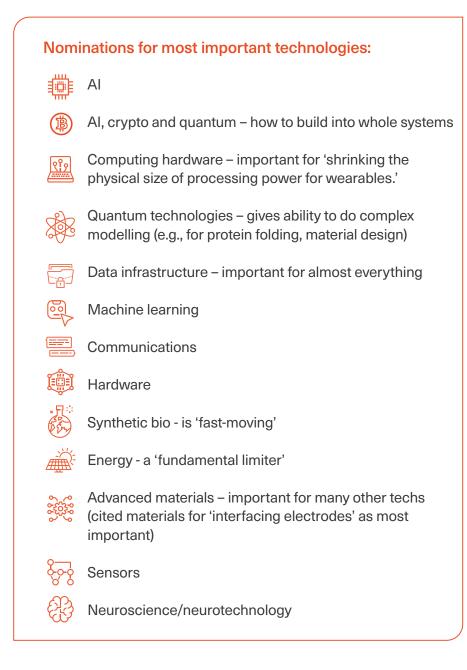


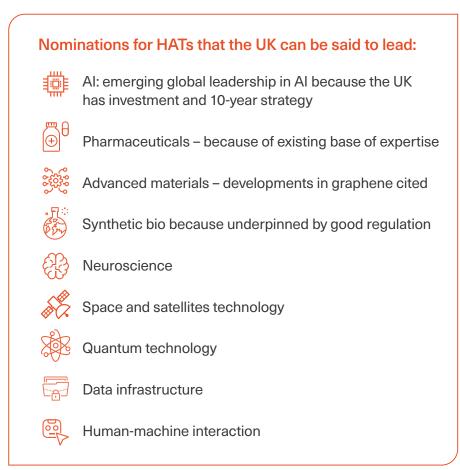
Session 2

From a prosperity and security perspective and with reference to the NS&I list of technologies, which HATs need to be protected in the UK?

Commentary on NS&I list:

- Impossible to compare or rank one above another
- Some in list are cross-cutting, enabling everything (e.g., everything is about data); others are niche
- The NS&I list should include printed electronics; power harvesting; and neurotechnology





Are HA applications always sensitive for UK national security?

- Depends on definition of 'national security'
- No depends on context, use and individual cases.
 (e.g., Google collecting information on where you parked your car is less sensitive than personalized health data)
- There are different orders of national security concern
- Human error/user error are bigger risks than malicious attacks
- Supplies that make up certain components of tech solutions which are made in foreign territories (China was cited) cannot always be trusted
- The potential ability to change the way people think or what they think with HAT is very sensitive (e.g., intentional mass attacks – 'social media injected directly into brain'; 'streaming content into eye implant'; and, in the theatre of war, 'changing the minds of commanders')



How should the UK balance security considerations with economic/academic openness to successfully innovate while protecting its world-leading IP?

Note: In the following sections, tables infer an association between contextual issues and stated solutions and principles. It is not known whether participants intended these associations.

Context	Solutions and principles
UK is seen to be bad at exploiting IP, perhaps because the gap between research and market is so difficult to bridge	Create longer term national programmes which support the translation of technologies - the National Quantum Technologies Programme is a good example of this
Academics' need to publish openly clashes with security concerns when working on sensitive research	Determine purpose of university research departments – research or teaching?
Difficult to keep innovative research within national boundaries; governments can't prevent companies from selling tech Governments not big enough customers for innovators to get a return on investment	Leverage valuable UK-based IP to access IP from around the world Have open standards for security – does not impinge on IP
Funders do not take into account security level when providing funding	Have national security guidelines for companies developing devices by uniting with international partners. For example, funders should not 'pass go' on any funding application without national security considerations being alerted. Research Councils and Ethics Panels for consideration. Ensure devices are 'secure by design'
Restraint for security reasons blocks R&D progress. (e.g., rules restricting IP use outside of Government use restricts progress because it restricts international collaboration)	Defence and security accelerator model for R&D/Government- sponsored incubator labs – protects IP, encourages exploitation Establish more mature framework for civil defence operations
Success in R&D innovation often necessitates protecting IP	Open-source libraries show that there are ways around this
UK is seen as risk-averse and so tends to lose in technology races	
Science research is international and collaborative, but regulation is insular/about protecting home interests	
Risk that big breakthroughs will happen in control of non-state actors, many of whom are lead HATs	



Session 3

What principles could ensure these technologies are used ethically and safely, for the positive benefit of humanity?

Short-term impacts for technology developers to manage:

Issues	Solutions and principles
Not knowing long-term impacts; lack of long-term evidence of safety/ efficacy Verifying that tech delivers on its	Protocol for risk assessment/ validation (what are consequences, and could we live with them?)
promise/does what it aims to do	Scrutinize impacts on early adopters as part of experimental phase. Formalize long-term follow-up to monitor safety/efficacy over long term
Ethics not embedded in tech (e.g., Al) Tech with negative effects that are not reversible Unethical eugenics, especially involving heritable characteristics Life in metaverse very different – difficult to adapt, learn	Learn lessons from other developments in other sectors – e.g., AI – and tackle those first Retrofit ethical principles in AI Incentivize ethically sound objectives/actions so that they are more attractive than unsound ones Make tech and effects of tech reversible, if possible Restrict level of genome modification
Controlling direction and scope of overall R&D effort Bar to implementation of tech is getting lower, costs getting cheaper, opening it up to malicious actors or abuse	
Implantable devices for non-medical uses (Neurolink) potentially beyond safety net of regulation Tech that falls between gaps in regulations and standards	



Long-term impacts for technology developers to manage:

Issues	Solutions and principles
Generating and maintaining public trust	Be transparent in objectives, explain the tech, and seek consent
Protecting privacy Lots of different ethical frameworks	Encourage co-creation, especially between developer disciplines and
Lots of different ethical frameworks confuse the picture	governors Co-development with public input to make compact with society and secure their consent.
	Consider ways that tech can be abused, including by infringing privacy (e.g., detecting identities from brain signals)
	Standardized internationally agreed framework for assessing safety and ethical risk – generic framework for application across all technologies and applications
	Establish mechanisms for encouraging positive behaviours and discouraging negative ones
	Close gulf between developers' and governors' motivations and understanding
	Account for cultural factors
	Consider benefits in the round – i.e., balance benefits against the tech's potentially stigmatising effects (e.g., of being a cyberhuman)
Ongoing safety monitoring ('legacy of devices') might be beyond clinical support	Development of international standards
	Retain accountability and responsibility for whole life of product/service – foresee difficult scenarios, such as supplier going bust
	Consider whole life of tech and whether it can be reversed
	Assign accountability and responsibility throughout whole life of devices
	Extend responsibility for after-market users to protect early adopters

Long-term impacts for technology developers to manage (cont):

Issues	Solutions and principles
Risk aversion hinders R&D Developers not planning far enough into the future	Prioritize and fund fully/properly to attract best people and allow best R&D processes
Technology (until Al gains consciousness) only gains ethical dimension when applied/used	Shift R&D focus from NS&I list to as-yet unknown technologies, with scenario-testing possible impacts
	Test future scenarios 20 years in advance to anticipate possible future applications
	Ensure all long-term considerations are also part of short-term ones
	Have different ethical frameworks for researchers (pure science) and developers (exploitation impacts citizens)
Application of HATs gets political because it entails unethical security	Amend human rights laws to include augmented human rights.
risks to the augmented individual Impacts vary at different scales:	Consider both positive and negative risks – at all scales
what is small for the individual might be huge for society (e.g., in climate change debate, environmental impacts)	Build in a failsafe that can stop development if negatives clearly outweigh positives ('big red button')
	Consider widest possible user groups (e.g., tech developed for disabled user group might end up being used by people outside of that group)



Short-term impacts for technology governors to manage:

Issues	Solutions and principles
Tech raises political, mental, social and ethical issues	Ensure diverse/inclusive representation and technical quality assurance
Attitudes shift over time, cultures and geographies Differing applications of tech have different contexts, resulting in differing standards Lack of interest in and harmonization with other nations' and jurisdictions' regulations, which influences what is developed	 Develop international standards around devices, and ensure they are flexible to deal with future developments Speed up regulatory pathways Regulatory pathways should be transparent Simplify and harmonize the regulatory landscape/codes of conduct to guide developers so that they engage (avoid 'shoulder-shrugging mentality') Policy statements about how we want to collaborate with foreign interests would set the compass
Knowing what data should be collected and what should not, and what to do with useful but not consented information	



Long-term impacts for technology governors to manage:

Issues	Solutions and principles
Lack of control means developers continue to conduct R&D despite uncertainty, potential ethical issues, or in ignorance of ethical issues Short-term benefit might lead to long- term harm for patient's health and their personal life Limited governing control over ownership of data and who gets to exploit data Limited governing control of ownership of and rights to tech Limited consideration of social inequality impact – do enhancements affect class structures? Dignity of augmented people should be preserved	Embed ethical principles into regulation and standards – be prepared Provide ethical oversight during development Apply the concept of what we might regret in the future, and to what extent, to set guidelines Require post-market monitoring and planning (e.g., to protect augmented human if company goes bust or ignores signals)
Adhering to standards is prohibitively expensive for small companies	Accelerate regulated pathways
Challenge is adoption of new tech by clinicians NHS is diverse and dispersed organization, making it harder for it to act as a unified client of scale	Mechanisms for easing adoption in NHS – build accessibility from ground up
Governance drives in a single direction, and can stifle innovation	Develop R&D communities with lots of independent voices
Not understanding how devices or parts of devices work is a security risk Unsecured communications channels (Bluetooth) are a security risk	A system for verifying security of devices in regulation Global governing body to identify security threats – agile enough to keep up?
R&D must be rewarded with resources (money), and so we need to understand where the resources/ money comes from Commercial acquisitions are complex, expensive and almost impossible to track, let alone regulate. (Cited example of UK Government scrutiny of NVIDIA's attempted takeover of Arm)	Regulate funding to mitigate risk of malign influence Regulate and police commercial acquisitions to prevent too much power accumulating in single companies (It is noted that this is what NS&I attempts to do, and the NVIDIA/Arm case demonstrates that regulation is possible. Of course, sphere of influence is restricted to British context, though)





Connecting for Positive Change

Innovate UK KTN helps businesses get the best out of creativity, ideas and the latest discoveries, to strengthen the UK economy and improve people's lives.

As a network partner of Innovate UK, Innovate UK KTN links new ideas and opportunities with expertise, markets and finance through our network of businesses, universities, funders and investors. From materials to energy and from manufacturing to healthcare, Innovate UK KTN combines in-depth knowledge in all sectors with the ability to cross boundaries.

Connecting with Innovate UK KTN can lead you to potential partners, horizon-expanding events, bespoke support and innovation insights relevant to your needs.

Innovate UK KTN Suite 218 Business Design Centre 52 Upper Street Islington London N1 0QH

03333 403251 enquiries@ktn-uk.org ktn-uk.org @KTNUK