AI + Emerging Technologies and future Cybercrime

Prof. Philip Treleaven
University College London
p.treleaven@ucl.ac.uk
www.financialcomputing.org

Major challenge is ‘Awareness’

- AI and Emerging Technologies (DeepTech)
- Cybercrime innovations (CrimeTech)
- Information Security Technology (InfoSec)
- Law Enforcement (LawTech)

My History (co-founded 9 start-ups)

Research

- **UK National Sizing Survey**
  - 11,000 adults scanned for 14 major UK retailers

Government Advisor

- **Japan MITI Fifth Generation Project**
- **European Commission ESPRIT programme**
- **Korea, Singapore, Thailand, ...**
- **Luxembourg, Malta ... FinTech ecosystem**
- **Bank of England, Financial Conduct Authority, ...**

Cyberspace –

environment in which communication occurs over computer networks

- **Cybercrime – criminal activities carried out using computers or internet.**
  - **Devices-as-Targets:** phishing, DOS ransomware.
  - **Devices-as-Tools:** deepfakes, romance scams.
- **Dark practices – User Interface carefully crafted to trick users**
  - **Dark patterns:** utilities 'harvesting' customer bank accounts, overpriced insurance etc.
  - **Dark content**: misinformation, dark web, deviant content etc.
  - **Digital addition:** encouraging addictive use of an online application.
- **Institutions – governments, organisations ... monitor, control or influence.**
  - **Social 'norm':** promoting a social or political agenda, cyber bullying.
  - **Surveillance**: security services, China Social Credit System, etc.
  - **Subversion**: (foreign) governments or organisations seeking to influence or impact society or markets.
  - **Predictive 'policing'**: for the future, identifying potential 'illegal' activity.
- **Algorithms – rise of nefarious algorithm either deliberately created or evolving.**
  - **Feral algorithms:** evolve unintentionally to corrupt/conspire with Superintelligence algorithms.
  - **Criminal algorithms:** deliberately created criminal algorithms.

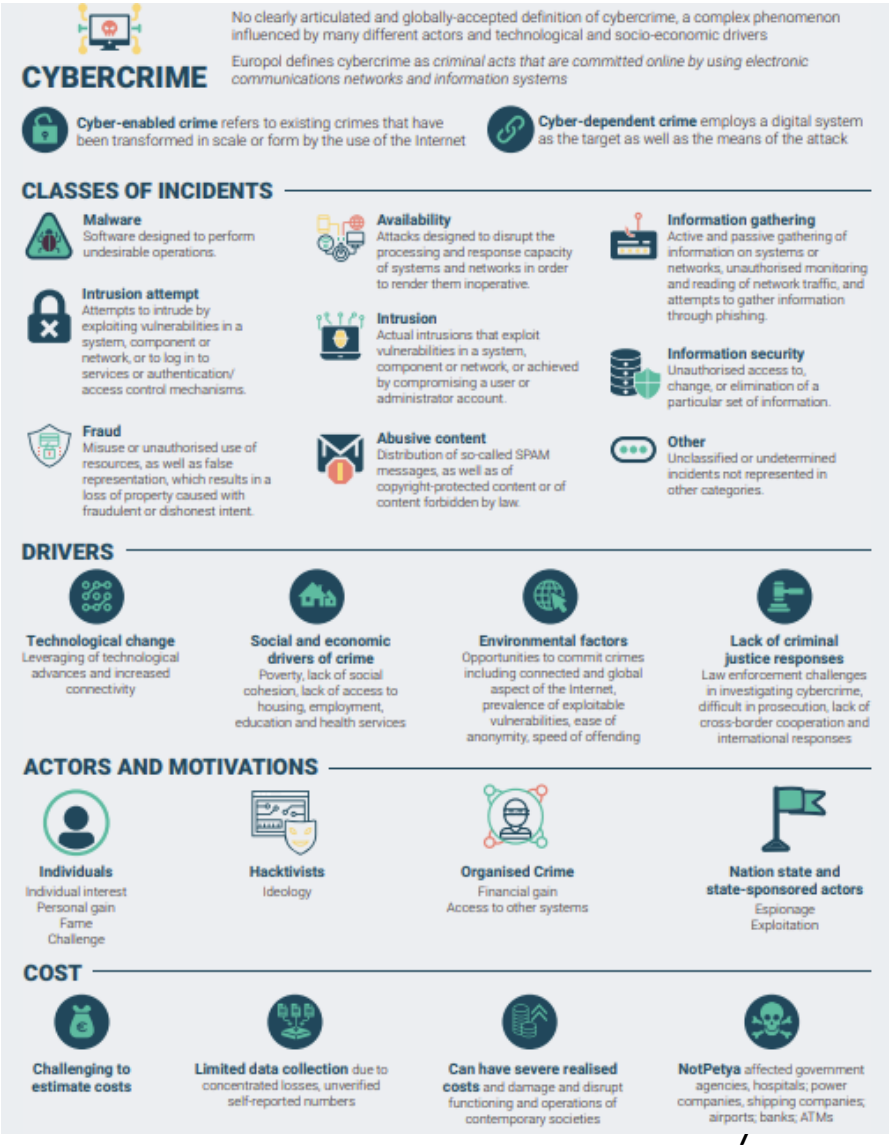
Globalization of Cybercrime and Law Enforcement

	Traditional	Future
Agents	▪ Humans, Institutions	▪ Algorithms, avatars, androids, DAOs
Analytics	▪ Retrospective analysis	▪ Real-time analysis
Information	▪ Data	▪ NLP text, images, speech, video
Interventions	▪ Identified cybercrime activities	▪ Previous unseen patters of cybercrime
Jurisdictions	▪ National, identifiable Agents	▪ Global, anonymous Agents
Networked devices	▪ Smart phones, computers	▪ Smart devices, IoT infrastructure devices

Cybercrime *industrialization*

Cybercrime *ecosystem*

- **Sophistication** - cybercriminals are becoming more collaborative, specialized and using Generative AI for programs.
- **Collaboration** - between state-sponsored operatives and cybercriminals is on the rise (e.g., Iran, Russia and North Korea).
- **Cyberattacks** - on critical data and IoT infrastructure, including ransomware, and becoming more expensive.
- **Botnets** - automated malware deployment tools are becoming more sophisticated by utilizing AI and emerging technologies.
- **Threat** – all organizations of all sizes are in danger, but especially SMEs who frequently lack sophisticated resources for anomaly detection.



Criminal Algorithms *- new actors*

- **Actors – anonymous humans, algorithms, avatars, androids, companies**
- **Criminal algorithms – algorithms, avatars and androids deliberately created or evolving to commit crimes.**
- **Algorithm anomalies – real-time detection and prevention: identifying patterns in algorithm behavior that do not conform to well-defined notions.**
- **Algorithm authentication – the process of proving or verifying algorithms and avatars.**
- **Algorithm conduct – need for new class of machine learning algorithms with a built-in understanding of bias, ethics, fairness, risk, and legality etc.**

Cybercrime *– a perfect storm*

- **Cyber extortion – users accessing dark content, or the dark web, open themselves to threats of reputational damage, until a ransom is paid.**
- **Cyberterrorism - the politically motivated use of computers and information technology to cause severe disruption or widespread fear in society.**
- **Dark content – anonymous agents, deepfakes and the metaverse will see an escalation of illegal deviant content. Opening users to cyber extortion.**
- **Deepfakes - deep learning used to create convincing texts, images, speech, and video hoaxes, resulting in bogus content and fakes.**
- **Digital addition –encouraging excessive and harmful behaviors, using technologies such as recommender systems in retail and tokenization in 'play-to-earn' gaming.**
- **Economic volatility – market risk and crises due to unpredictable or deliberate algorithm behaviour (cf. flash crash 2010).**
- **Misinformation – inaccurate or deliberately misleading information leading to algorithmic bias and misinformation caused by bad or deliberately biased training data.**
- **Social manipulation – social and political manipulation through AI algorithms and misinformation information.**
- **Social surveillance – surveillance with emerging technology, such as China's Social Credit System.**
- **Weaponization – autonomous weapons androids powered by AI.**

AI + Emerging Technologies

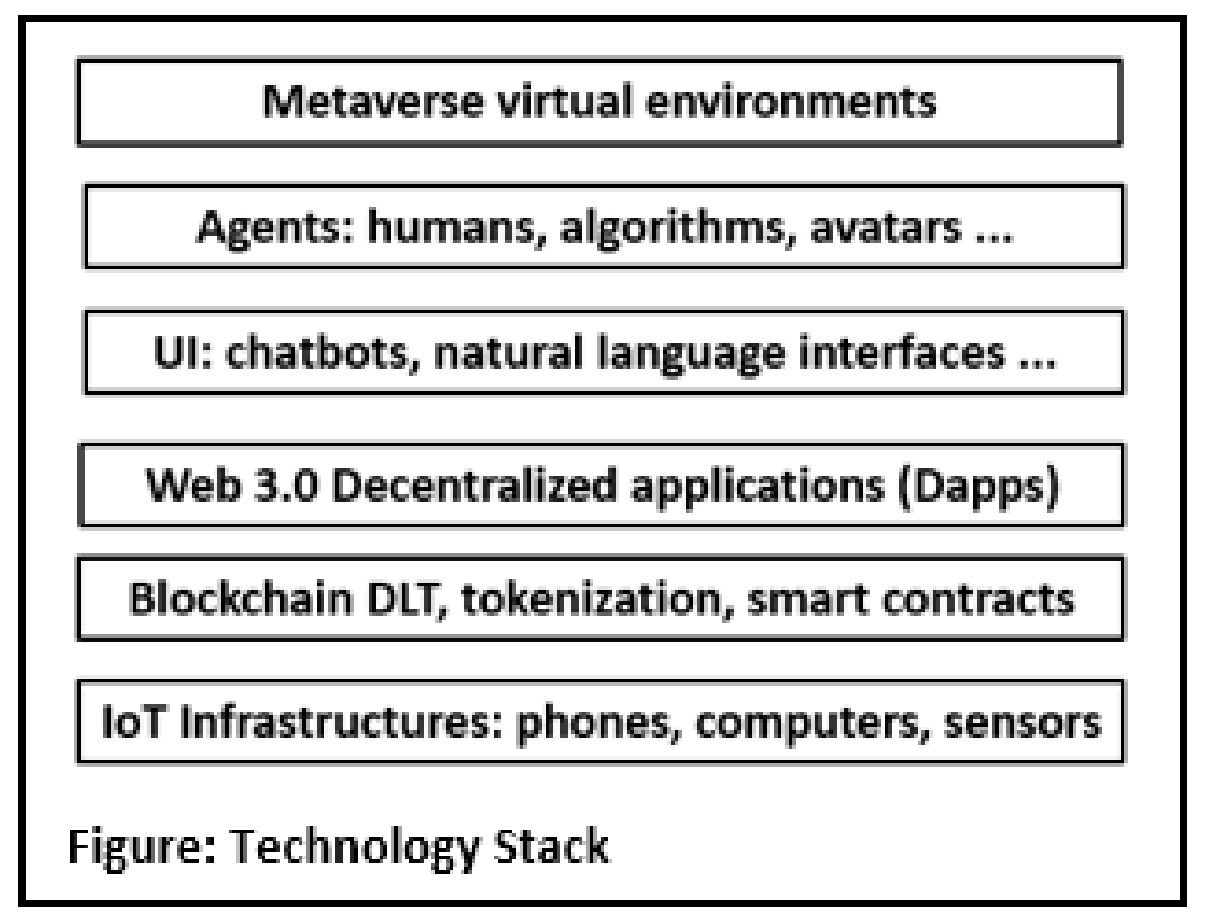


Figure: Technology Stack

AI + Emerging Technologies

- **Artificial Intelligence**
  - **Agents** – now humans, algorithms, avatars, androids, and organizations; often anonymous and global.
  - **Algorithms** – traditional machine learning systems for automation, now generative AI for user interaction, and in future Algorithmic Superintelligence.
- **Emerging Technologies**
  - **User Interface (UI)** – chatbots, natural language interaction (e.g., text, images, speech, video), avatars, VR/AR and metaverse virtual environments etc.
  - **Blockchain DLT** – using distributed ledger, cryptocurrencies, tokenization, and smart contracts.
  - **Decentralized environments** – Web 3.0 decentralized application systems (i.e., Dapps) and IoT infrastructures of autonomous smartphones, computers, and devices in vehicles, machines, and buildings etc.

AI and Humans working in partnership

'human-in-the-loop'

- **Algorithmic Intelligence – algorithms are good at focused data analysis and automation but can mis-interpret results, *hallucinating* and are poor at long-term vision and forecasting.**

- **Human Intelligence – professionals are frequently good at broad insights and vision, such as risk and reputation.**

- **AI Superintelligence - an AI system that is self-aware and intelligent enough to surpass the cognitive abilities of humans.**
  - **Feral algorithms** – that evolve into cybercriminals and corrupt or conspire with other Superintelligence algorithms.
  - **Opaque algorithms** – where the behaviour is unpredictable and uninterpretable.
  - **Virus-like algorithms** – deliberately created algorithms that replicate in uncontrollable ways.

AI algorithms generations

- **Knowledge-based algorithms** - rule-based systems where knowledge is explicitly represented as ontologies or IF-THEN rules rather than implicitly via code.
- **Machine Learning** - able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data.
  - **Traditional ML** - machine learning (ML) for automation: data scraping, analytics, transacting (e.g., algorithmic trading).
  - **Generative AI** - – large language model (LLM) algorithms (such as ChatGPT) to create 'human-like' content, including audio, images, text, videos, even code and simulations.
  - **Artificial General Intelligence (AGI)** - machine intelligence that solve problems as well as a human.
  - **Artificial Superintelligence (ASI)** - machine intelligence that solve all problems better than people.
- **Brain–computer interface (BCI)** – Elon Musk's Neuralink, a direct communication pathway between the brain's electrical activity and an external device, smart phone, computer, or robot.

ChatGPT and Generative AI

- **Chatbot** - programs simulating conversation with humans: text, speech, images, video.
- **Transformers** - a neural network that learns context and thus meaning by tracking relationships in sequential data, like the words in this sentence.
- **Generative Pre-trained Transformers (GPT)** – AI (machine learning) models trained using internet data to generate any type of text, images, speech, video.
- **Large Language Models (LLM)** - LLMs are AI models trained on a vast quantity of [online] data to produce human-like responses to dialogue or other natural language inputs.
- **Generative AI** - a broad label that's used to describe any type of AI (e.g., neural networks) used to dynamically create new 'human-like' texts, images, speech, video, programs, or synthetic data.

Blockchain technologies

- **Distributed ledger** - an 'audit trail' of cryptographically encoded transactions.
- **Blockchain** - a distributed database that maintains a continuously growing list of ordered records, called blocks, linked using cryptography.
- **Cryptography** - a method of securing information and communications through the use of codes.
- **Cryptocurrencies** - a digital currency designed to work as a medium of exchange.
- **Tokenization** - the process of assigning a unique digital identifier and properties to an asset that's usable on a blockchain application.
- **Smart contract** – a program, representing an agreement, stored on a blockchain that automates the execution.

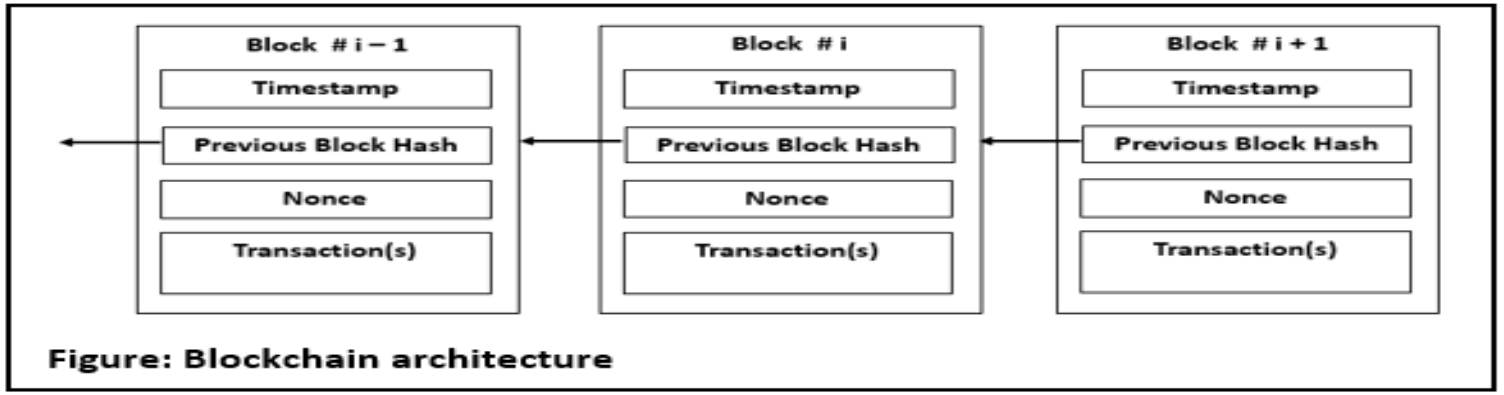
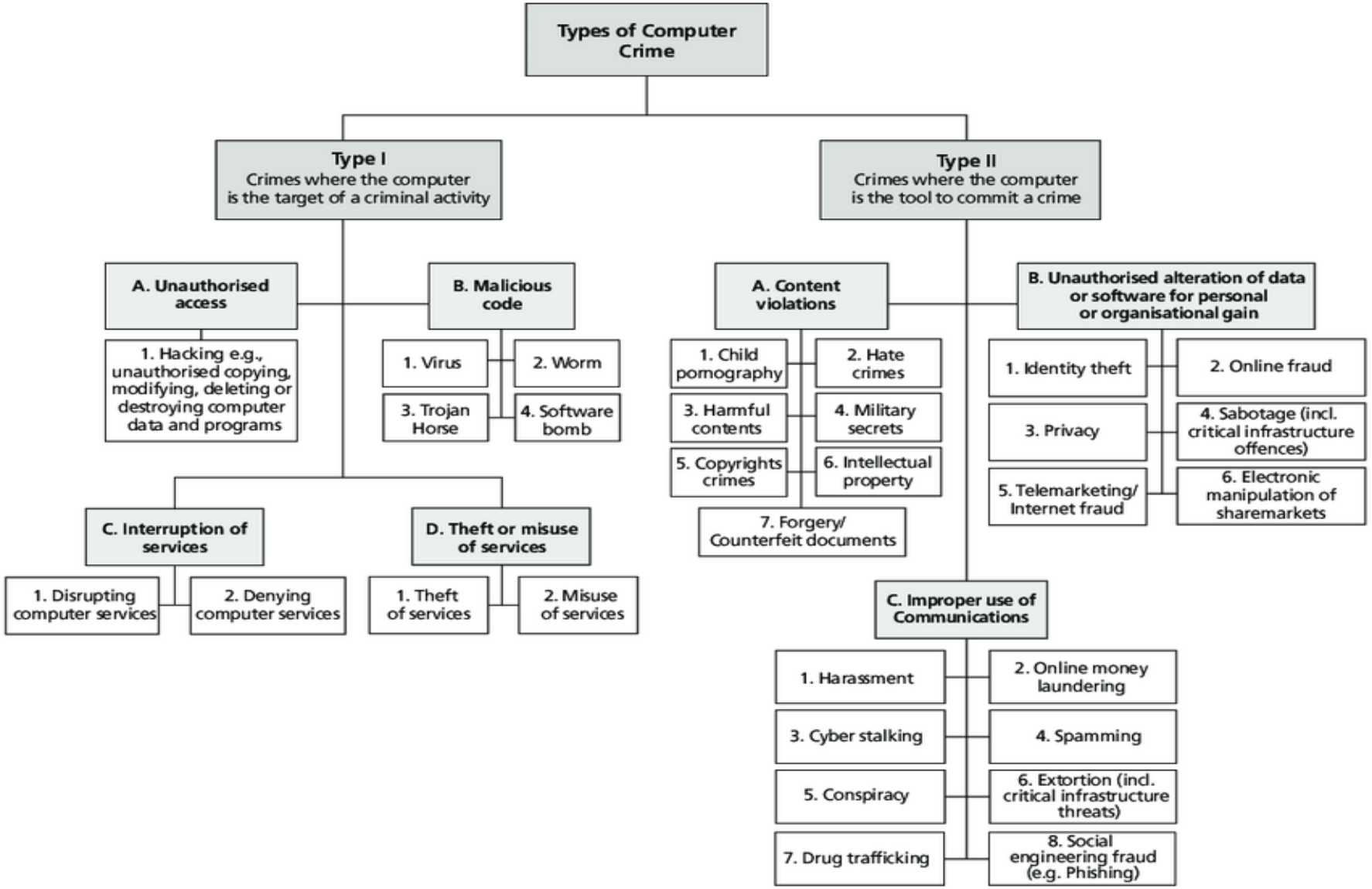


Figure: Blockchain architecture

Cyber ‘crime’ – illegal to deceptive

Cybercrime types



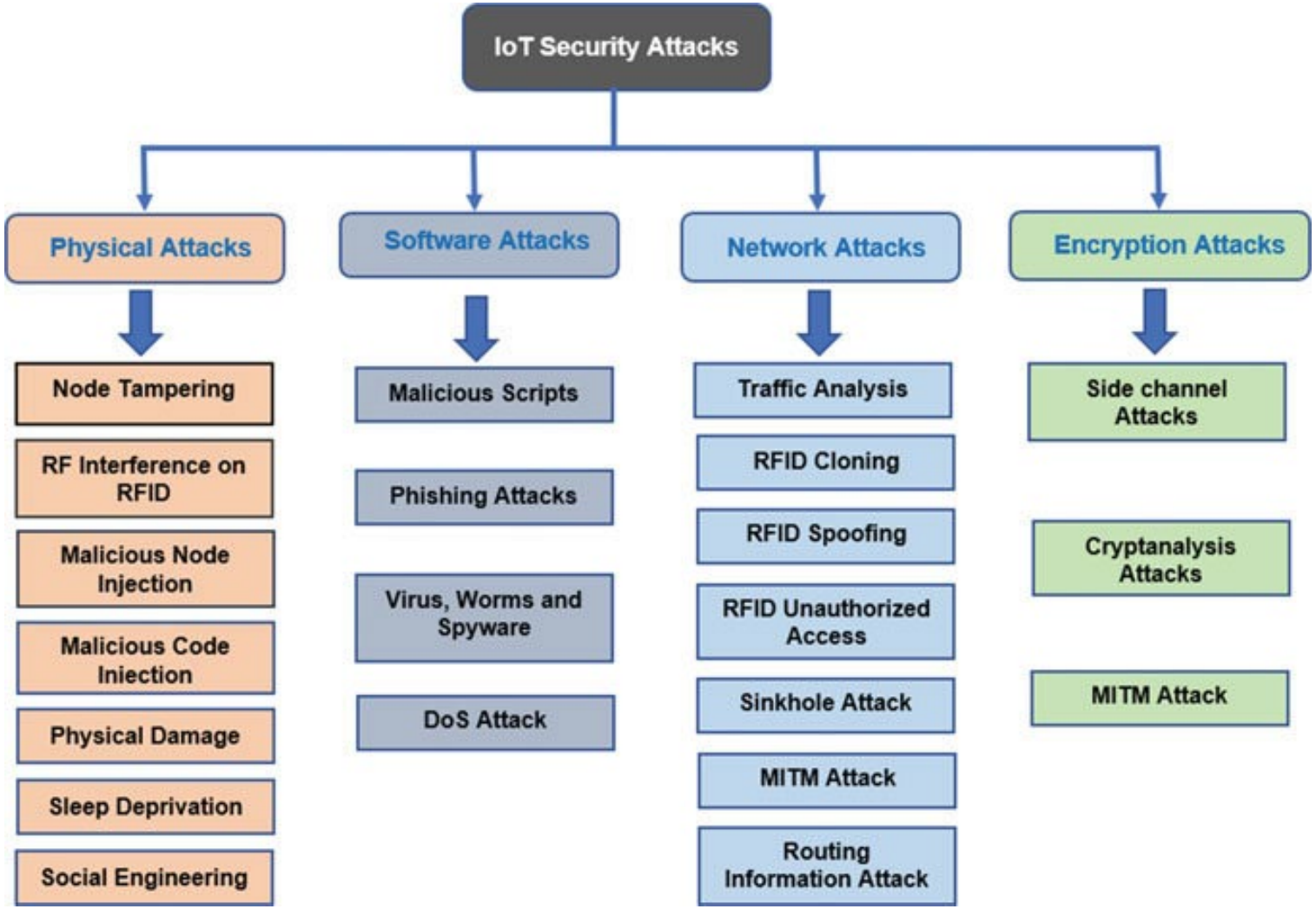
Devices-as-targets

- **Dark content – anonymous agents, deepfakes and the metaverse will see an escalation of illegal deviant content.**
- **Deepfakes – already generative AI can produce convincing bogus content and fakes. Applications range from influencers, financial ramping, reputational attacks, to cyber extortion.**
- **Digital addition – online games, betting sites and social media platforms are encouraging excessive and harmful behaviors, such as tokenization in 'play-to-earn' gaming.**
- **Denial of Service – control of servers and IoT infrastructure.**

Devices-as-tools

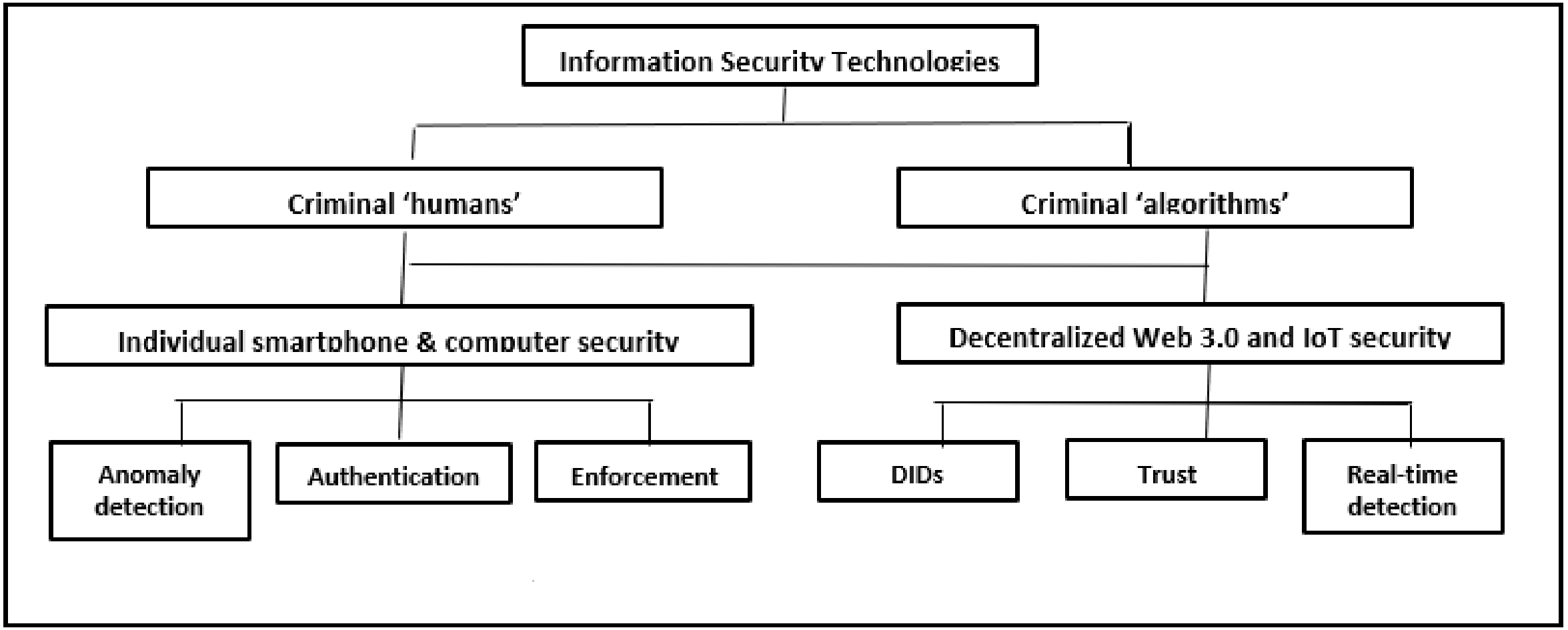
- **Cyber espionage – expect a major expansion of spying by corporations, governments, and criminal organizations (e.g., North Korea); and increasing collaboration of these players, both for financial gain and economic disruption.**
- **Cyber extortion - the use of various tactics, such as phishing, malware injection, and DDoS attacks to hold a victim's data or systems hostage until a (cryptocurrency) ransom is paid.**
- **Dark patterns – legitimate companies will increasingly 'trick' online customers through price inflation, while trying to avoid reputational damage.**
- **Dark web – expect a major expansion of the dark web (cf. parallel universe) with the rise of cybercrime, collusion of criminals with rogue governments, plus deviant metaverse sites.**
- **Data breach - the result of a hacker successfully breaking into a system, gaining control of its network, and exposing its data, usually personal data covering items such as credit card numbers, bank account numbers, Social Security numbers, and more.**
- **Spyware - increasingly governments will use Spyware to monitor opponents and companies for industrial espionage. Another form is Stalkerware, smartphone apps, and blue tooth hardware used for cyberstalking.**

IoT Infrastructures



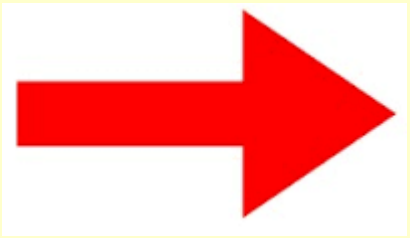
Information Security Technologies

InFoSec


Information Security Technologies
Criminal 'humans'
Criminal 'algorithms'
Individual smartphone & computer security
Decentralized Web 3.0 and IoT security
Anomaly detection
Authentication
Enforcement
DIDs
Trust
Real-time detection

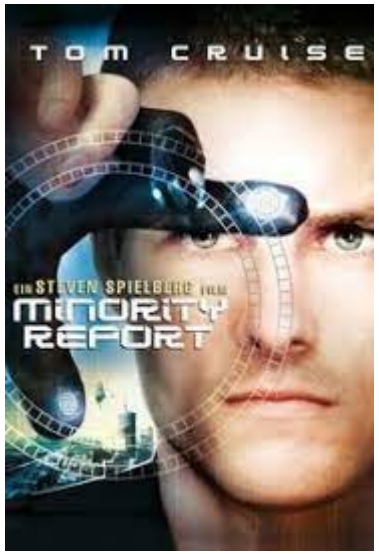
Law Enforcement *awareness*

balancing *innovation* and *regulation*



LawTech 3.0 *recommendations*

- **Horizon scanning** – detecting early signs of potentially important developments through a systematic examination of potential threats and opportunities, with emphasis on new technologies.
- **Knowledge-transfer** – raising awareness amongst stakeholders of CrimeTech and LawTech emerging technologies.
- **LawTech infrastructure** - to support secure networks for national and law enforcement collaboration together with (e.g., XML) information standards for coordination.
- **Predictive analytics** – pioneering the use of AI algorithms to analyze massive amounts of information in order to (even) predict and help prevent potential future cybercrimes.
- **Rapid-start laws** – to engage with Justice Ministries and politicians to provide legal provisions for rapid response to cybercrime 'innovations' (cf. Financial Regulators).
- **Sandboxes** – provide a LawTech testing environment where new or untested technologies and software can be run securely.
- **Self-reporting** – to engage with AI and emerging technologies' communities to encourage self-regulation and self-reporting of emerging cybercrime trends.
- **Tech sprints** – essentially hackathons, coding events that brings LawTech programmers and other interested people drive innovations.



Future-proofing your Ecosystem

AI Social/Business Revolution - empowering your staff

The genie is already out of the Bottle

Regulation and Law Enforcement *awareness*

- **CrimeTech awareness** – CrimeTech and LawTech training for all stakeholders (e.g., law enforcement, regulators, policy makers, politicians).
- **PoliceTech** – proactive policing, proactive enforcement, and for the future predictive policing systems dedicated to apprehending and detaining people before they have the opportunity to commit a crime.
- **JusTech** – automation of justice, such as recidivism algorithms, and algorithmic dispute resolution.
- **RegTech** – technology for automating compliance and regulation.

AI Social Revolution

(imagine if you couldn't use Microsoft Word, Email, Spreadsheet ...)

- **Boards** – institution awareness and strategy (e.g., Tesco)

- **AI Professional Tools** – Lawyers, Doctors, Programmers ... are using AI productivity tools as assistants to draft 'recommendations'. AI legal tools are drafting contracts, and identifying risks, allowing the human lawyer to refine the contract.

- **Training** - AI Chatbot Trainers will read a textbook and then be available to discuss the content with a staff member or student.

Staff need (re)education in the technology to understand the opportunities and weaknesses

Questions & Comments